The Impacts of Micro Stores on Labor Market Duality and Employability in Saudi Arabia: 1970–2020

Fahad Abdullah A. Aldubayan
M.A. in Economics, Southern Methodist University, USA
Lecturer of Economics, Al-Imam Mohammad Ibn Saud Islamic University, KSA

Thesis submitted in fulfilment of the requirements for the degree of
Doctor of Philosophy in Economics

Business School, Victoria University, Australia
May 2021

Abstract

In response to various economic challenges in the Saudi economy such as persistent unemployment, low labor participation, and poor performance of the private sector, both policymakers and researchers have endeavored to find ways of mitigating these problems. Researchers have focused their attention on three main theories: human capital theory, institutional theory, and socioeconomic theory. The first theory—human capital theory—is primarily concerned with labor supply. Researchers have used this theory to determine whether the Saudi labor force has sufficient skills as required by the private sector. The skill mismatch hypothesis has been used to explain the continuous high unemployment rates among Saudis; however, this may not be a convincing argument for several reasons discussed in this thesis. The second theory—institutional theory—focuses on localization policies. The broad argument is that although enforcing localization policies may succeed in integrating national labor into the private sector in the short term, this could lead to negative consequences in the long term. The final theory—socioeconomic theory—focuses on economic and social factors affecting the labor decisions of individuals (employers and employees). This thesis shows that these factors may not be the ultimate causes of high unemployment and low employability of Saudi labor in the private sector. What seems to be overlooked by these theories is the investigation of Saudi labor demand, and whether the private sector is capable of playing a significant role in creating meaningful jobs for the increasing labor force.

This thesis argues that three primary market-distorting factors have undermined the efficiency of the Saudi labor market, leading to a segmented labor market and ultimately, persistent unemployment. These factors stem from three policies, namely, (i) the rapid expansion of the public sector and attraction of Saudi labor, (ii) the rapid expansion of micro stores, and (iii) the sponsorship (Kafala) system. Together, these market-distorting factors have led to a structural shift, leading toward an inefficient labor market. In other words, the three policies have unintentional consequences that distort the market mechanism of labor demand and supply, thereby undermining economic efficiency. Overall, while the sponsorship system has distorted the total labor supply, micro stores have distorted the product markets by adopting a low capital-intensive model and shifting the labor demand toward low-skilled foreign workers. The public sector, for its part, with its appealing employment policies compared to the private sector, has resulted in a sectoral preference against the private sector. Consequently, it

can be considered that the Saudi labor market is segmented into primary (big-industry) and secondary (micro-industry) labor markets. Thus, by using the nested CES production function, this thesis estimates the direct and partial elasticities of substitution to examine Saudi and non-Saudi labor substitutability in these two labor markets.

The main findings can be summarized as follows. The range estimates of direct elasticity in micro and big industries are (-7.1; 16) and (-0.18; 0.55), respectively. The range estimates of partial elasticity in micro and big industries are (0.33; 0.77) and (0.51; 0.6), respectively. In both industries, the estimates of the direct elasticity of substitution have interval estimates containing zero; hence, it cannot be ruled out that the relationship between Saudi labor and capital is a perfect complementary relationship, which indicates the important role of capital investment in Saudi employment growth. In both industries, the estimates of the partial elasticity of substitution also have interval estimates between zero and one, providing evidence in favor of a complementary relationship between composite inputs (Saudi labor and capital) and foreign labor (i.e., they have a very low degree of substitution). These findings corroborate prior research that cautions against the counterproductive effects of localization (Saudization) policies. Two corrective measures are proposed in the thesis. First, it is suggested that the sponsorship system be replaced by a central recruitment agency. Second, it is suggested that micro stores be re-zoned or restricted within self-contained centers, and consortiums between the government and private investors be created to establish “transformative development companies” for developing certain industries and product markets while recruiting and training national labor.

Key Words: Saudization, Saudi unemployment, micro stores, unemployment persistence, labor market segmentation, dual labor market, expatriates, nested CES production function, (JEL Classification System: D2, E24, J08, J23, J42, J46, J68, K31, L25, O15, O17, O53)

Doctor of Philosophy Declaration

“I, Fahad Abdullah A. Aldubayan, declare that the PhD thesis entitled ‘The Impacts of Micro Stores on Labor Market Duality and Employability in Saudi Arabia: 1970–2020’ is no more than 80,000 words in length including quotes and exclusive of tables, figures, appendices, bibliography, references and footnotes. This thesis contains no material that has been submitted previously, in whole or in part, for the award of any other academic degree or diploma. Except where otherwise indicated, this thesis is my own work”.

“I have conducted my research in alignment with the Australian Code for the Responsible Conduct of Research and Victoria University’s Higher Degree by Research Policy and Procedures”.

Signature:  Date: May 25, 2021

Dedication

To my first teachers—my mother and my father.

With love …

Acknowledgments

First and foremost, the thesis would not have been possible without the blessings of Almighty Allah. All praise is due to Almighty Allah in this world and hereafter as Allah loves and is pleased with. Peace be upon Prophet Muhammad.

I would like to express my sincere thanks and deep appreciation to **Professor Tran Van Hoa** who has supported me throughout my candidature with constructive criticism and intellectual insight at each stage of my research. I am also deeply indebted to my associate supervisor, **Professor Sardar M. N. Islam**, for his invaluable suggestions and continuous support throughout my research journey. I have been truly fortunate to have worked with these two supervisors. I am thankful for the panel committee and my colleagues at the Victoria University Business School, who contributed to this thesis through their comments and suggestions. In particular, I am grateful to the panel chair, **Dr. Riccardo Natoli**, and other panel committee members, **Professor Jamie Doughney** and **Dr. Kim Sweeny**, for their invaluable suggestions and academic input at each milestone. I would also like to thank the Deputy Directors of the Institute of Sustainable Industries and Liveable Cities, **Professor Randall Robinson**, **Professor Dianne Hall**, and **Dr. Ricardo Natoli**, for their continuous support for researchers and for mitigating any challenges that may face researchers during their studies. Further, I acknowledge that my thesis was edited by Elite Editing, and editorial intervention was restricted to Standards D and E of the *Australian Standards for Editing Practice*.

I am most grateful to Imam Mohammad Ibn Saud Islamic University, Saudi Arabia, which provided the scholarship enabling me to pursue this PhD study.

Most importantly, I am indebted to my parents for their continuous prayers, their encouragement, and their forgiveness for my absence all these years. I also thank my wife, daughters, brothers, and sisters for their sacrifice and support at times when I need it most.

Table of Contents

List of Tables

List of Figures

List of Appendices

List of Abbreviations

CRA	Central Recruitment Agency
CES	Constant Elasticity of Substitution
CPI	Consumer Price Index
GASTAT	General Authority for Statistics (Saudi Arabia)
GDP	Gross Domestic Product
LD	Labor Demand
LM	Labor Market
LS	Labor Supply
SAMA	Saudi Arabian Monetary Agency
SABIC	Saudi Basic Industrial Corporation
SR	Saudi Riyal
ARAMCO	Arabian-American Oil Company
DEA	Differential Evolution Algorithm
GCC	Gulf Cooperation Council
ISIC	International Standard Industrial Classification
KSA	The Kingdom of Saudi Arabia
LMA	Levenberg-Marquardt Algorithm
UK	The United Kingdom
US/USA	The United States of America
VIF	Variance Inflation Factor
TDC	Transformative Development Companies
USD	United States Dollar

1. Introduction

1.1 A Brief Historical Review of Economic Development of the Kingdom of Saudi Arabia

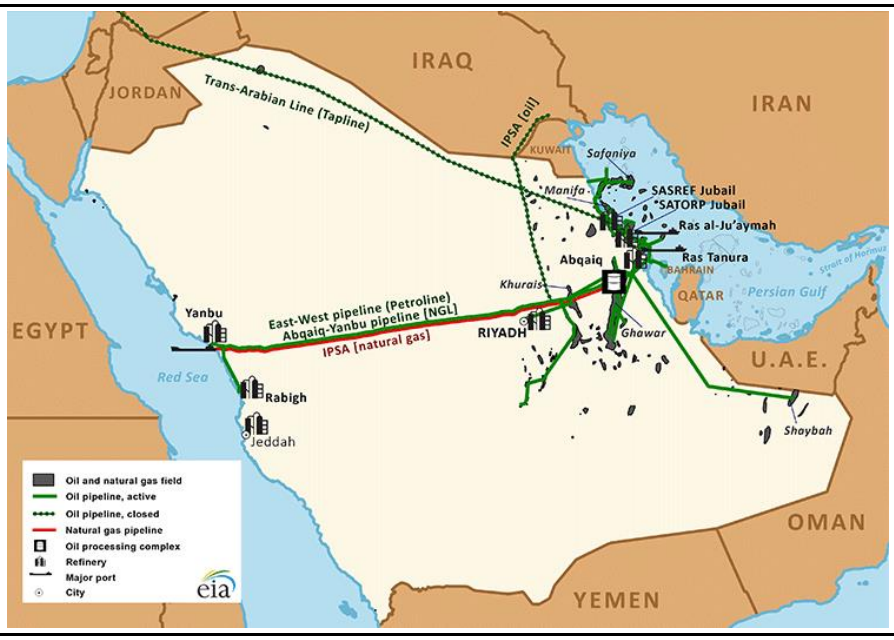
1902–1970

In 1902, Riyadh, a small town at the heart of the Arabian Peninsula, was conquered by Abdulaziz Ibn Saud. He continued capturing other towns and extending the country's boundaries over the majority of the Arabian Peninsula until 1932, when he founded what is known today as the Kingdom of Saudi Arabia (KSA). Ibn Saud went on to became the first king of the third (current) Saudi State (Citino, 2002). In the early twentieth century, most of the population of the Saudi State were Bedouins, scattered all over the Arabian Peninsula, engaged in a continuous quest for food and water. The Bedouins were generous, patient, loyal and tribal people. But they were also illiterate and uncontrollable since they did not settle in one place. Therefore, in 1912, King Abdulaziz instituted an agrarian policy by founding small colonies (*Hijrahs*)[1] to encourage the Bedouins to settle. This would allow him to send the *Ikhwan*[2] to teach the Bedouins the basic principles of Islam, thereby binding them to fight beside the king during times of war (Citino, 2002; Rihani, 1954). Nowadays, although there are no *Ikhwan*, many colonies established at that time still exist today, although, like many other towns in Saudi Arabia, they are lagging behind in terms of modern development. The lack of modern development in many towns may work as a "push factor," prompting rural–urban migration, partly explaining the low mobility of Saudi labor between different cities.

[1] In Arabic, one colony is called Hijrah, which is derived from the Arabic word al-Hijrah (i.e., migration) since the Bedouins migrated from the desert to settle in colonies.
[2] The tribesmen who fought with King Abdulaziz. They were known for their religious ardor and strong influence on their tribes.

Before the development of the oil industry, Saudi society was largely traditional. The main source of livelihood was pastoral agriculture and trade while the main source of government revenue was taxes generated from hundreds of thousands of annual pilgrims to Mecca (Niblock & Malik, 2007; Sirageldin et al., 1984). Although oil was discovered in 1938, its large-scale production only commenced after the second world war (Edens & Snavely, 1970). During the post-war period, the first oil industry development occurred in Eastern province, when the Tapline project (see Figure 1.1) was constructed between 1947 and 1950. In 1950, for geopolitical and economic reasons, the US government arranged a deal between King Abdulaziz and the Arabian–American Oil Company (ARAMCO) with equal profit sharing of oil revenues. ARAMCO was allowed to deduct the Saudi share from corporate taxes since the US government desired to support ARAMCO in continuing its mission of discovering more oil fields in the KSA (Citino, 2002). As a result, the KSA became a rentier state, dependent on the revenue of depleting natural resources. Almost all Saudi government revenues at the time were generated by the oil industry, allowing the government to spend more on defense and invest in US Treasury Bonds. Although there was a slight increase in incoming foreign labor, the economy was still primitive, with the primary driver being handicraft (Vassiliev, 2000).

Figure 1.1: Saudi Arabia Major Oil and Gas Infrastructure



Source: (U.S. Energy Information Administration, 2020)

In the late 1950s, after a relative increase in oil prices and unplanned government expenditure, the Saudi government suffered from a financial crisis resulting from the fiscal deficit and currency depreciation. The crisis was severe enough for the state to seek assistance from the International Monetary Fund (Niblock & Malik, 2007). Consequently, in 1962, Prince Faisal, the Head of the Ministries Council at the time, sanctioned the “Ten-Point Program,” which can be considered the first attempt at economic development in Saudi economic history. Because this program was not fully implemented for various reasons, the economic and social conditions continued to be unsatisfactory. For instance, there were relative increases in income inequality and poverty as well as a high level of illiteracy among the population (Edens & Snavely, 1970; Vassiliev, 2000). Consequently, the government felt compelled to improve the economic and social conditions of the population. As a second attempt to build the base of the nation’s infrastructure and avoid any fiscal deficits such as those experienced in the late 1950s, Prince Faisal decreed establishing the Central Planning Office in 1965. The office was mandated to draft five-year economic plans (quinquennial plans). Thus, in 1968, the Central Planning Office announced its first Five-Year Development Plan, set to start in 1970.

1970–2015

The decade of the 1970s was a turning point in the economic history of the KSA. The high revenue generated from the surge in oil prices enabled the Saudi government to deliver what it had promised in the first three Five-Year Development Plans (1970–1984). This boom resulted in the large-scale transformation of both the economy and society. Economic development began with the main cities, such as Riyadh and Jeddah, and several towns in the Eastern provenance, making them “pull factors” for rural–urban migration. Because of urbanization efforts, the Saudi government succeeded in decreasing the nomadic population from about 42% in 1966 to just 7% in 1992 (Mahdi, 2000). However, rural–urban migration created imbalances in the supply and demand of labor in cities, worsening labor immobility, and resulting in labor

shortages. While there was a demand for workers to take over jobs in relatively abandoned rural areas, there was an even larger labor demand in recently developed cities because of the large scale of new projects and expansion of the public sector. Sirageldin et al. (1984) point out that in the 1970s, the era of capital constraint on development ended, replaced by a new era of labor constraint. This hindered development, leading to a serious labor dilemma. Consequently, the Saudi government allowed the influx of foreign labor as a short-term solution, which is discussed further in section 3.2.3.

Between 1970 and 2014, the Saudi economy oscillated between expansionary and contractionary fiscal policies, reflecting fluctuations in oil price. Generally, expansionary fiscal policy was adopted over the periods between 1970–1984 and 2005–2014 while contractionary fiscal policy was adopted over the period between 1985–2004. In the expansionary periods, Saudi employment increased as a result of jobs created by new public projects in various sectors. In the contractionary period, Saudi employment also increased, though at a slower pace, because policymakers embarked on recruitment in the public sector as a necessary measure to counter economic and political challenges. Further, during the period of contractionary fiscal policy, Saudi Arabia, as a rentier state dependent on oil revenues, experienced economic and institutional inertia. For instance, after oil prices plummeted in 1983, the economic conditions in Saudi Arabia were characterized by five factors: “(i) lower growth rates in many sectors and a decline in others; (ii) intense competition; (iii) a tightened lending policy on the part of banks and development funds; (iv) the disappearance of abnormal profits; and (v) greater fluctuation of exchange rates” (Al Hajjar & Presley, 1996, p. 108).

Moreover, the drop in oil revenues made the Five-Year Development Plans ineffective while the private sector lacked the ability and/or incentives to continue economic development (Wilson et al., 2012). Lastly, some of the ramifications of the 1990s financial crisis stemmed from the substantial reduction in oil revenues, including:

> the relatively low level of foreign assets and reserves, the rise of unemployment, the continuing and substantial deficit in the balance of payments, the inadequate facilities available for the population as a result of too little public investment in the social and economic infrastructure, and the worsening of the poverty. (Niblock & Malik, 2007, p. 174)

Therefore, the limited job growth in the private sector on the one hand, and the limited capacity of the public sector to recruit an increasing labor force on the other, may have contributed to high and persistent unemployment among Saudis, particularly women and the young.

Over the period between 2005 and 2014, because of the substantial financial reserves generated from high oil prices at the time and because of the lack of projects undertaken across the country for the last 20 years (during the contractionary fiscal period), the government felt compelled to resume its economic development agenda. Thus, new projects were embarked on in various sectors such as health, education, and even building so-called Economic Cities from scratch. Although recruitment in the public sector resumed at a higher pace, unemployment rates remained high. The eagerness of the government to mitigate unemployment and income inequalities resulted in an increase in employment in the public sector, eventually leading to underemployment and low productivity.

In short, the Saudi economy suffered from low levels of productivity and a lack of resilience in the face of oil price fluctuations. While the state is still largely dependent on oil revenues, the private sector is critically dependent on government support (purchases and subsidies). Hence, because of the absence of countercyclical policies (Wilson et al., 2012), the private sector is always susceptible to drops in oil price. Such an economic state necessitates urgent intervention; however, the government has been seemingly reluctant to do so because of the economic and social burdens that usually accompany reform[3].

[3] Much of the questions and arguments of this research were developed before the announcement of the recent economic plan, Saudi Vision 2030. Because this plan is still in the process of being implemented, it may be too early for its economic evaluation, and discussing it would take us beyond the scope of this research.

Hence, it seems there is no clear “exit mechanism” out of the current dilemma of the Saudi economy while maintaining market efficiency. Over the last 50 years, the private sector has shown itself to be incapable or apathetic to take the initiative and create valuable jobs despite continuous support from the government. At the same time, the size of government has increased because of undertaking development projects by itself, resulting in negative impacts on economic performance. In the next section, it is argued that three market-distorting factors have undermined the efficiency of the Saudi labor market, leading to a segmented market and, in turn, persistent unemployment as well as other labor issues.

1.2 Background of the Research Problem

As mentioned earlier, the decade of the 1970s was a turning point in the economic history of the KSA. This period saw the government implementing its Five-Year Development Plans. Various policies were embarked on that transformed the economy and the labor market structure, some of which had unintended consequences. The study identifies three primary policies that had unintended consequences by distorting the Saudi labor market, leading to a segmented market, and therefore, persisting unemployment among local labor. These policies are (i) the rapid expansion of the public sector and the attraction of Saudi labor, (ii) the rapid expansion of micro stores, and (iii) the sponsorship (*Kafala*) system. These policies are termed in this thesis as the “market-distorting factors,” and are discussed thoroughly in Chapter 3 while they are explained here briefly as follows.

The first policy was the rapid expansion in the public sector and the attraction of Saudi labor resulting from the increase in government expenditure at an unprecedented rate. Implementing the Five-Year Development Plans, particularly the first three plans (1970–1985), at a rapid pace, resulted in a sudden increase in public goods and services. For instance, the basic

infrastructure of the country, such as interstate roads, public schools, ministries, ports, airports, hospitals, new cities, and universities were built within a short time. This resulted in an urgent need for the government to attract Saudi workers by offering better incentives than the private sector offered, such as higher wages, regular working hours, training, and the possibility of promotion. This led to a tight labor market situation, placing upward pressure on Saudi wages. The private sector was, and still is, largely undeveloped because of the dominance of the informal sector, impeding the private sector from competing with the public sector. Consequently, most Saudi labor abandoned the private sector in favor of the public sector. Since then, the public sector has become the first-and-last resort employer of Saudi labor, while the private sector is critically dependent on foreign labor.

The second policy was the rapid growth of micro stores, resulting from urban planning and the expansion of cities. A micro store can be defined as any shop employing less than five workers. It can be described as an informal business, depending primarily on cheap low-skilled foreign labor, and having exceedingly low levels of capital expenditure, which negatively affects technological diffusion and working conditions.[4] This is a business model that has widely emerged since the 1970s because of the expansion of cities and urban planning, which allowed an excessive number of micro stores to proliferate. Micro stores can be established along so-called “commercial streets,” together with central markets (*Souqs*). Unlike the idea of a “high street,” where certain areas are designated for commercial purposes, a “commercial street” in Saudi Arabia could be virtually any street in the city grid, resulting in an excessive number of micro stores as well as other environmental and security issues. Consequently, the policy of land use, which allows micro stores to be built along the “commercial streets” as opposed, for example, to designated and carefully planned shopping centers, has contributed, as this study argues, to redundant micro stores, and thereby, to a *disguised* shortage of national labor as well as to the emergence of the informal labor market. In addition, the increasing number of micro stores has gradually

[4] See some pictures of micro stores in Appendix 1 for better visualization of micro stores in Saudi Arabia.

reduced the importance and profits of central markets, thereby displacing many Saudi workers who used to work there. It should be mentioned at this point that, throughout this study, two business models are contrasted: micro stores and professional/formal firms. The former were defined above; the latter can be described as having a combination of all the qualities linked to good working conditions. The good working conditions primarily consist of high wages and internal administrative rules determining employment relations, such as recruiting and training workers for specific skills, which typically exist in large firms. In this study, “professional” and “formal” are used interchangeably to describe firms with such qualities. In addition, this research uses "meaningful jobs" terminology. There is no consensus on the definition of "meaningful jobs". In this research, however, meaningful jobs are those "good" jobs that are created by professional/formal firms and offer high wages and fringe benefits by which employees would have a sense of relatedness and job security. Also, in this research, “good” jobs are contrasted with "bad" jobs that created by micro stores which are in bad working conditions and do not offer high wages nor opportunities for job advancement to attract qualified Saudi labor.

These two policies—pertaining to the rapid expansion of the public sector and the micro stores— required a large labor force to build and operate the expanding size of government and the growing number of micro stores. However, these policies were challenged by the shortage of Saudi workers in terms of both quantity and quality (skills) needed in the public and private sectors. This was because of the rapid expansion of the public sector and the micro stores, which were growing at higher rates than the Saudi labor force. For instance, between 1988 and 2016, the average annual growth rate of new shop licenses was around 7%[5] (Ministry of Municipal and Rural Affairs, 2018). In contrast, the annual average growth rate of the total Saudi population for the same period was around 2% (Saudi Arabian Monetary Authority, 2016). To meet such labor shortages, the government initially sought to attract Saudi workers using higher wages and

[5] Despite the unavailability of data prior to 1988, it is likely that the growth rates of shops were higher than is stated here. This is because the greatest expansion was in the 1970s, subsequently abating due to the economic recession and contraction in the fiscal policy.

fringe benefits. However, the labor shortage continued, particularly in labor-intensive departments, such as education and health services. The attraction of Saudi labor to the public sector has resulted in a crowding-out effect, reducing the Saudi labor supply available to other sectors in the economy. The substantial government expenditure and attraction of Saudi workers have resulted in crowding-out effects in the private sector. This is because the government became the main provider of many goods and services as well as the main employer of Saudi workers at that time.

At the time, two approaches were considered by the Saudi government to overcome the shortage of the labor in both the public and private sectors (Sirageldin et al., 1984). The first was to take a gradual approach and build the nation using the local labor force; the other was to take a more rapid approach and import the needed foreign workers. The substantial oil revenues and the urgent need for major infrastructure projects may have been the main reasons for adopting the second approach. Consequently, the third policy came into place. This involved endorsing a short-term policy by relaxing entry restrictions for guest workers into the Saudi labor market under temporary work visas. This came to be known as the sponsorship (*Kafala*) system (Al-Thaqafi, 2000; Sirageldin et al., 1984). The sponsorship system facilitated the influx of foreigners into the Saudi labor market even though most foreign workers were redundant because of the excessive number of micro stores, which has disguised the shortage of local labor. As a result, the ratio of foreign workers to the total population increased from 7% in 1963, to 31% in 1980 (Sherbiny, 1984b), and to 37% in 2016 (Saudi Arabian Monetary Authority, 2016). In 2010, foreign labor constituted about 77% of total labor in the Saudi private sector (General Authority for Statistics, 2010a).

The critical difference between the public and private sectors when it comes to the national labor shortage is that in the public sector, the main need for foreign workers is mainly to build, but not to operate (i.e., a temporary need), while in the private sector, the main need was to build *and* operate (i.e., a continuous need). Unlike the public sector, the private sector is arguably incapable of

recruiting Saudi labor because it is a largely informal sector dominated by micro stores. Hence, the private sector is experiencing a continuous dependency on cheap unskilled foreign labor. Even with labor-intensive departments in the public sector, the Saudi government managed to meet the labor shortage in the public sector by attracting Saudi workers with high wages and fringe benefits and training them to fit their new jobs. Thus, while the labor shortage in the public sector was genuine and overcome by attracting Saudi labor, the labor shortage in the private sector is continuous and mostly generated *artificially* by increasing redundant micro stores.

In addition, since the public sector is non-profit, the government has succeeded, to a greater extent compared to the private sector, in replacing foreign workers. In contrast, the private sector, as a profit-seeking sector dominated by micro stores, is not eager to replace cheap foreign laborers with relatively more expensive Saudi laborers. The influx of foreign labor is facilitated by the Sponsor System, which allows foreigners to renew their work visa annually regardless of their residency period in the country. In other words, they are always treated as guest workers, and they cannot become permanent residents. At the same time, Saudi workers are not eager to work in micro stores for various reasons. One prominent reason is that micro store jobs offer wages considered lower than the subsistence level given the standard of living in Saudi society. Because of the low start-up costs of micro stores, Saudi workers could also establish their own micro stores instead of working for others. If they do not have enough money to establish micro stores, other means, such as family and government support, are usually easily accessible. The preference of some Saudis to set up their own business rather than working for others increases the proliferation of micro stores. This phenomenon could be explained by the model of firm size distribution developed by (Lucas, 1978), as explained in section 3.2.3. Micro store owners, like any business owners, are entitled to apply for work visas on behalf of foreign workers, assuming they would come as employees, not freelancers. It turns out that many micro store owners sublease their micro stores to foreign workers to generate a passive income (i.e., rent-seeking activities), a

practice known as “commercial concealment,” which is discussed further in section 3.2.3.

In short, as the number of micro stores increases, so does the number of foreign workers. The rising number of foreign workers results in, among other effects, reduced wage levels and low participation rates of local labor in the private sector. Therefore, this study argues that the micro store model cannot be compared with the small enterprises, which prevail in other countries and play an important role in creating job opportunities (discussed further in section 3.2.2). While there is a labor market duality between the public and private sectors, it can be argued that micro stores have contributed to a dual labor market within the private sector, with foreign laborers working in micro stores (the informal/secondary labor market) and Saudi laborers working in large professional firms (the formal/primary labor market).[6] Consequently, the Saudi government is facing a substantial challenge in attempting to reduce unemployment by enforcing localization (“Saudization”) policies because of the limited jobs in the primary labor market and undesirable jobs in the secondary labor market. Indeed, the Saudi economy is facing a unique phenomenon, necessitating the careful investigation of the “micro stores effect” in conjunction with other market-distorting factors. These factors are fundamental to understand the labor market segmentation and the high, persistent unemployment among the local labor force in Saudi Arabia.

[6] This study uses “formal” and “primary” as well as “informal” and “secondary” interchangeably.

1.3 Objectives and Structure of the Research

The primary objectives of this study are to examine the employability of Saudi labor as well as investigating the conditions of the Saudi labor market. These objectives can be addressed through the following questions:

- Why is unemployment high and persistent among Saudi labor?
- What is the role of the public and private sectors in employment growth?
- What are the market-distorting factors that could be the fundamental reasons for unemployment persistence over the last several decades despite the government's generous support to education and businesses to increase Saudi employment in the private sector?
- How has the micro stores model, as one of the market-distorting factors, contributed to the so-called dual labor market? And how does it differ from the small and medium enterprises?
- What should be done to reverse the effects of the market-distorting factors and correct market failure, thereby creating meaningful job opportunities for the increasing local labor force?
- Does the elasticity of substitution between Saudi and foreign workers in micro industries differ from that in big industries?
- What are the possible implications of the elasticity of substitution on localization policies and Saudi employment?

The study is organized into six chapters. Chapter 1 is the introductory chapter and consists of six sections that present (i) a brief historical review of the economic development of the KSA, (ii) a background of the research problem, (iii) the objectives and the structure of this thesis, (iv) the limitations of prior studies, (v) the methodology used in this study, and (vi) the theoretical and practical contributions of this study.

Chapter 2 places the labor issue into perspective through an extensive literature review consisting of four sections: (i) an overview of the characteristics of the Saudi labor market, (ii) a brief description of the different economic schools

of thought regarding unemployment (iii) a discussion of three theories used to explain high unemployment in Saudi Arabia, namely, the human capital, the institutional, and socioeconomic theories, and (iv) an outline of the limitations of previous studies to identify the gap in literature on local unemployment in the KSA.

After discussing the limitations of previous studies investigating Saudi unemployment, Chapter 3 builds an alternative economic framework to obtain a better understanding of the unemployment issue. The chapter is organized into three sections. (i) The first section consists of the following three sub-sections discussing the market-distorting factors: (a) the rapid expansion of the public sector and attraction of Saudi labor, (b) the rapid expansion of micro stores and the disguised shortages of the national workforce, (c) the sponsorship system. (ii) The second section illustrates the effects of market-distorting factors and unemployment using graphical analysis. As the thesis argues, these market-distorting factors may have caused structural shifts in the Saudi labor market, leading to labor market duality. This duality is clearly evident between the public and private sector and less clearly evident within the private sector, between the formal and informal labor markets. It is argued that this market segmentation has ultimately led to persistently high unemployment among the local labor force. (iii) The third section examines the role that government can play to reverse the effects of the market-distorting factors and reduce market failure, thereby generating meaningful job opportunities for the increasing local labor force.

Chapter 4 is devoted to empirical analysis and is organized into two main sections on (i) the econometric model and research hypotheses and (ii) data sources and estimation method. The nested constant elasticity of substitution (CES) production function is used to estimate the direct elasticity between Saudi labor and capital, and to estimate the partial elasticity between Saudi labor and capital on the one hand and foreign labor in the other, using the data of big and micro establishments. Chapter 5 presents the results and discusses the implications of the study. It is organized into three sections: (i) descriptive statistics and variance inflation factor, (ii) parameter estimates and the elasticity

of substitution including the hypothesis testing, and (iii) the implications of the study. Last, Chapter 6 concludes the thesis.

1.4 Limitations of Prior Studies

Previous studies on the Saudi labor market and unemployment have relative importance and potential implications for the Saudi economy. However, they are constrained by one or more of the following limitations.

First, the studies lack consistent economic analysis explaining why unemployment persists in the country despite the government's generous support for education and business to increase Saudi employment in the private sector. It is also despite the improvement of skills attainment among the Saudi labor recent decades. Hence, it can be argued that previous studies do not *fully* explain persistent unemployment. In contrast, this study identifies three market-distorting factors that have contributed to a dual labor market within the private sector and perpetuate unemployment among the local labor force.

Second, focusing only on labor supply theories (e.g., skill mismatch hypothesis), while neglecting labor demand (e.g., market structure and working conditions) has been proven to be an inadequate approach, given the persistent high educated unemployment and the other labor issues in the Saudi economy. Although the skill mismatch hypothesis might explain a portion of the unemployment, it cannot explain the persistent, high unemployment for various reasons (discussed in section 2.4.1). Focusing solely on labor supply theories ignores insights derived from other theories. For instance, previous studies lack implementing theories of labor market segmentation, such as the dual labor market theory. This theory seems consistent with the situation of the Saudi labor market insofar as it explains both the low national labor participation in the private sector as well as the persistence of the high unemployment rate.

Third, prior studies lack an estimation of the CES between Saudi labor, capital, and foreign labor considering different establishment sizes. The present study uses the nested CES production functions to estimate the elasticities of substitution to evaluate the substitutability between three important production inputs: Saudi labor, capital, and foreign labor, considering different establishment sizes, which might provide empirical evidence to assess the localization policy.

Finally, prior research fails to investigate the effects of micro stores on (un)employment despite the prevalence of this model in the Saudi labor market. In addition, these studies have virtually failed to consider the unemployment in the Saudi labor force from the industrial organization perspective despite the dramatic changes to the market structure over the past decades. Having identified the main limitations of prior studies, the discussion now turns to the methodology used in this study and then to its main theoretical and practical contributions.

1.5 Methodology of the Study and Main Findings

In economics, research is typically conducted using normativism or positivism. The latter is used when researchers analyze “what is” questions, while the former is used when researchers focus on “what should be” questions, which is widely used in problem-solving and subject matter economic research (Ehrenberg & Smith, 2012; Ethridge, 2004). According to Angrist and Krueger (1999), 80% of papers published in labor economics contain some empirical analysis while two-thirds use micro data. The authors state that in the 1970s, the use of micro data became more popular than time series data. By the mid-1990s, micro data was used about ten times more frequently than time series data. Further, Angrist and Krueger (1999) distinguish between two differing yet complementary methods of empirical research in labor economics, namely, descriptive analysis and causal inference. Descriptive analysis considers facts about the labor market that need to be addressed by theoretical reasoning. The

authors emphasize the importance of descriptive analysis as an essential step that comes before theorizing because it provides facts used to build on theories. In contrast, causal inference identifies the effects of specific policy interventions or estimates behavioral relationships.

In this study, descriptive analysis is used in Chapter 2 to establish facts about the Saudi labor market. In Chapter 3, a theoretical framework is proposed about the impacts of micro stores and other market-distorting factors. In the empirical analysis, the nested CES production function is used to estimate the direct elasticity of substitution between Saudi labor and capital, and the partial elasticity of substitution between composite inputs (Saudi labor and capital) and foreign labor in both micro and big establishments. To the best of the researcher's knowledge, this is the first attempt in the context of the Saudi labor market, where the nested CES production function is used to provide new insight and a deeper understanding of the Saudi labor market. This is achieved by using secondary data on the Saudi market structure to determine whether the elasticity of substitution among different production inputs differs in two different labor markets. Particularly, the Establishments Economic Survey is used. This survey is published annually by the General Authority for Statistics (GASTAT) and based on the Enumerating Establishments. The Enumerating Establishments was the first-ever census of establishments, published in 2010 by General Authority for Statistics (2010a). It covers the number of establishments, Saudi workers, non-Saudi workers, compensations, expenditure, value added and gross capital formation. All variables are disaggregated according to three categories of establishment sizes, classified by the number of employees (less than 5, 5 to 19, and 20 or more), except for "gross capital formation," which is reported in total (aggregated). The data are time series, consisting of 83 cross-sectional industries, listed according to the International Standard Industrial Classification (ISIC4), and eight years covering the period 2010–2017 (General Authority for Statistics, 2017a). Therefore, there are 83 sub-industrial sectors and eight years, yielding a total of 664 observations. The other details of the methodology are discussed in Chapter 4.

It is crucial to pay close attention to the substitutability relationship between Saudi and non-Saudi labor in different labor markets since this has critical implications for the analysis. In the primary labor market, the relationship between Saudi and non-Saudi labor can be substitutable or can have complementary relationships according to the type of skills or occupations. For example, if Saudi and non-Saudi laborers have similar skills (such as middle-level jobs), the relationship would be substitutable. Hence, replacing foreign workers would be possible because Saudi workers are willingness to take on such jobs. However, if foreign workers have specific skills that Saudi workers do not possess (such as upper-end jobs) or if they hold certain jobs that Saudi workers do not usually accept (such as lower-end jobs), the relationship would be complementary.

In contrast, in the secondary labor market, the relationship between Saudi and non-Saudi laborers might be mostly complementary because of the undesirable characteristics in this type of labor market. It is also considered a complementary relationship because the micro stores model resembles a passive income model, rather than an entrepreneurial one, whereby Saudi workers play the role of employers (sponsors) and foreign workers play the role of employees (renters). In fact, if such a relationship exists, the relationship between Saudi and non-Saudi laborers might become a perfect complementary one, as in the Leontief production function. Hence, the replacement mechanism would be expected to fail when the relationship between Saudi and non-Saudi laborers were a complementary relationship. Thus, enforcing the localization policy might be counterproductive, resulting in serious ramifications.

Consequently, it can be considered that the Saudi labor market is segmented into primary (big-industry) and secondary (micro-industry) labor markets. Thus, this thesis estimates the direct elasticity of substitution (between Saudi labor and capital) and the partial elasticity of substitution (between the composite inputs and foreign labor) to examine the substitutability among different production inputs in these two labor markets. The empirical findings can be summarized as follows. The range estimates of direct elasticity in micro and big industries are (-

7.1; 16) and (-0.18; 0.55), respectively. The range estimates of partial elasticity in micro and big industries are (0.33; 0.77) and (0.51; 0.6), respectively. In both industries, the estimates of the direct elasticity of substitution have interval estimates containing zero; hence, it cannot be ruled out that the relationship between Saudi labor and capital is a perfect complementary relationship, which indicates the important role of capital investment in Saudi employment growth. In both industries, the estimates of the partial elasticity of substitution also have interval estimates between zero and one, providing evidence in favor of a complementary relationship between composite inputs (Saudi labor and capital) and foreign labor. These findings corroborate prior research that cautions against the counterproductive effects of localization policies.

In terms of descriptive statistics, the main features of the Saudi labor market can be summarized as follows. First, the structure of the Saudi industry and Saudi employment distribution reveals that the least localized (i.e., least Saudized) industries are the least concentrated while the most localized (i.e., most Saudized) industries are the most concentrated (see Table 2.8 and Table 2.9). Second, the Saudi market structure is dominated by a high number of fragmented establishments (i.e., micro stores and small establishments constitute 97% of establishments), which is likely to impede the transformation of the private sector into a developed, efficient sector (see Table 2.10). Third, the median estimates clearly reveal the difference in output between the micro and big industries (see Table 5.1). This is because the added value in big industries is about 16 times more than that in micro industries, despite the fact that the firm size distributions for micro and big industries are about 83% and 3%, respectively. In terms of the number of workers, big industries, by definition, have more labor (Saudi and non-Saudi). However, establishments in both micro and big industries recruit more non-Saudi workers than Saudi workers. Lastly, the net capital formation is aggregated; hence, we cannot comment on the size of each industry separately. However, the median value of the net capital formation is about SR645,000 only, which reveals that the capital investment in the private sector is low.

Two corrective measures are proposed in the thesis. First, it is suggested that the sponsorship system be replaced by a central recruitment agency. Second, it is suggested that micro stores be re-zoned or restricted within self-contained centers, and consortiums between the government and private investors be created to establish “transformative development companies” for developing certain industries and product markets while recruiting and training national labor.

1.6 Theoretical and Practical Contributions of the Study

This study has theoretical and practical implications. The theoretical implications can be summarized as follows.

1) Theorizing and estimating the impacts of micro stores may be considered the first academic endeavor of its type, despite the dominance of these stores in the Saudi economy since the 1970s.

2) Bringing the dual labor market theory into the analysis of Saudi unemployment may provide us with a better understanding of the Saudi market structure and lead to several theoretical implications.

3) Identifying the main factors that have unintentionally distorted, since the 1970s, the structure of the labor market of the KSA, may provide a more coherent explanation of persistent, high local unemployment. Researchers may continue, using the same or similar theoretical paradigm, to investigate the difference of the elasticity of substitution in different industry sizes instead of using overall estimation, which may not necessarily lead to a nuanced understanding of the substitution relationships between Saudi and non-Saudi labor in the Saudi market.

4) Unlike mainstream economists and policymakers, who advocate small-business models as a solution for Saudi unemployment, this study suggests that the status quo of the micro store model is a distorting model of the labor market and may hinder the performance of the private sector, thereby increasing the unemployment of the local workforce in the KSA.

5) This study uses the Establishments Economic Survey that is annually published by GASTAT and based on the Enumerating Establishments. This approach is adopted because, to the best of the researcher's knowledge, until date no study has used the Enumerating Establishments census despite its paramount importance in enhancing the understanding of the Saudi labor market structure. This study uses the 2010 census for the descriptive analysis and the survey series (2010–2017) for the empirical analysis. The Enumerating Establishments was the first-ever census of establishments, published in 2010 by General Authority for Statistics (2010a).

6) Using the nested CES production function enables us to estimate the direct elasticity of substitution between Saudi labor and capital, and the partial elasticity of substitution between composite inputs (Saudi labor and capital) and foreign labor as well as understanding the different relationships between those inputs in a dual labor market setting.

In terms of practical implications, there are several reasons for investigating the impacts of market-distorting factors, which can be summarized as follows.

First, the fact that both the rate of micro store jobs and local unemployment has increased in recent decades leads us to question whether the micro stores model generates meaningful job opportunities for the increasing local workforce. Investigating micro stores may draw attention to this economic phenomenon, thereby providing different policy implications for localization policy, training, and fostering different business models.

Second, recent estimates indicate that approximately 12% of the total local workforce is unemployed and that 74% and 54% of this unemployed workforce are in the 15–29 and 20–29 age groups, respectively (General Authority for Statistics, 2019a). Given the fact that the young Saudi population comprises a large portion of the total population, this may intensify the problem of unemployment and deepen despair among the Saudi youth, to the extent of threatening the stability of the country. This concern has been raised by other authors, especially after the so-called “Arab Spring” (Leber, 2019).

Third, the Saudi government aims to reduce its dependence on foreign workers as well as the oil industry by developing other industries and moving toward a knowledge-based economy. This is stated in several Five-Year Development Plans and in the most recent economic reform, Vision 2030 (Saudi Vision 2030, 2016). Thus, investigating micro stores focuses attention on their negative role in the labor market, which may hinder the achievement of such goals, especially given that they are the dominant business model in the Saudi economy.

Fourth, bringing theories of labor market segmentation into the analysis of the labor market assists us to evaluate different public policies, such as spending on education and training, subsidies for the private sector, and localization policies. For example, according to the dual labor market theory, the return on human capital in the secondary labor market is less than in the primary labor market (Doeringer & Piore, 1970; Osterman, 1975; Piore, 1968).

Fifth, technological progress has been facilitating a shift toward capital-intensive production and automation, which may result in a further increase in the unemployment rate. In addition, although boosting international trade opens up different opportunities for both producers and consumers, it may increase unemployment even further when firms move to other countries outsource operations abroad.

Finally, the Saudi government appears to be aware of its inability to continue as the first-and-last-resort employer for the increasing local labor force. Therefore, in recent years, the Saudi government has aimed to increase Saudi employment in the private sector by enforcing localization policies to replace foreign workers with Saudi workers and by introducing an expatriate levy. However, the localization policies could lead to counterproductive effects, especially in the long run (discussed in section 2.4.2.2). Moreover, by using the nested CES production functions, this study seeks to predict the impact of increasing the cost of foreign workers in Saudi employment in both big and micro establishments.

All in all, the market distorting factors are important in analyzing the Saudi labor market because they help explain at least seven central issues: (i) the high and persistent unemployment rate, (ii) the low labor participation rate among Saudi men and women or the slow process of integrating local labor into the private sector, (iii) the low human capital investment in specific skills through on-job-training programs such as internships and apprenticeships, (iv) the poor economic performance of the private sector and why it is susceptible to localization policies and any changes in government programs (subsidies and purchases), (v) the underemployment and low labor productivity in the public sector and its low resilience in the face of economic shocks resulting from fluctuations in oil prices, (vi) the high dependence on foreign labor in all industries, and (vii) the slow process of economic transformation and diversification.

2. Literature Review of Labor Economics

2.1 Introduction

This chapter reviews the literature on the Saudi labor market and unemployment theories. First, an overview of the characteristics of the Saudi labor market is presented, followed by a discussion of the different economic schools of thought addressing unemployment. Then, various reasons for Saudi unemployment are examined under three main sub-sections: the human capital, the institutional, and the socioeconomic theories. Finally, the limitations of previous studies are considered to identify the gap in literature on local unemployment in the KSA.

2.2 Characteristics of the Saudi Labor Market

In this section, an overview of the Saudi market structure sheds light on the size of each industry, their average percentage contributions to the gross domestic product (GDP), and distributions of establishment size in each industry as well as in the overall economy. This section also covers demographic aspects including the labor force and its characteristics. The review of the structure and distinguishing features of the economy provides a clearer perspective for the discussion that follows.

Table 2.1 lists the average annual growth rates of the sectoral contributions to the GDP from 1970 to 2017 at constant prices (2010=100).

Table 2.1: Average Annual Growth Rates of the Sectoral Contributions to the GDP from 1970 to 2017 at Constant Prices (2010=100)

Industries	Average
A.	
1) Agriculture, Forestry & Fishing	2.25%
2) Mining & Quarrying	56.15%
a) Crude Petroleum & Natural Gas	55.92%
b) Other	0.23%
3) Manufacturing	7.44%
a) Petroleum Refining	3.49%
b) Other	3.95%
4) Electricity, Gas and Water	0.67%
5) Construction	4.17%
6) Wholesale & Retail Trade, Restaurants & Hotels	4.63%
7) Transport, Storage & Communication	2.65%
8) Finance, Insurance, Real Estate & Business Services	7.60%
a) Ownership of Dwellings	4.64%
b) Others	2.96%
9) Community, Social & Personal Services	1.61%
10) Imputed Bank Services Charge	0.85%
Subtotal	86.32%
B.	
Producers of Government Services	12.94%
Total Except Import Duties	99.26%
Import Duties	0.74%
Gross Domestic Product	100.00%

Source: The average means were calculated by the researcher based on the GDP series (1970–2017) published by the General Authority for Statistics (2017c)

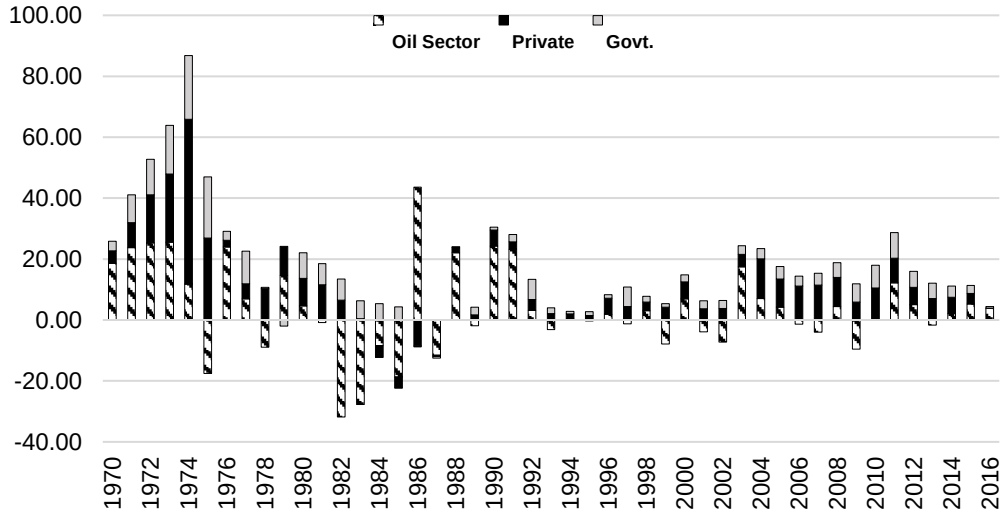
Clearly, the mining and quarrying industry stands out among all other industries as the largest industry contributor to the GDP with 56.15%. The industry has two sub-categories: (i) crude petroleum and natural gas, and (ii) “other,” with percentage contributions of 55.92% and 0.23%, respectively. Thus, the first category takes credit for contributing most to the GDP. Although this highlights the importance of the crude petroleum and natural gas industry, it equally shows the degree to which the Saudi economy is dependent on the oil industry and susceptible to risk exposure resulting from oil price volatility. The third and eighth industrial groups (manufacturing; finance, insurance, real estate

and business services) are ranked as the second and third largest contributors to the GDP with 7.44% and 7.60%, respectively. The rest of the other industrial contributions are less than 5% (except for producers of government services).

There are two points worth mentioning in Table 2.1. First, the industrial categories are slightly broader than the ISIC4 used in Enumerating Establishments Surveys, making the comparison somewhat inexact. The second and most important point is that this list may be slightly deceptive because of the heavy dependence of other industries on the oil industry. For instance, the manufacturing industry has two sub-categories: petroleum refining and “other,” with average percentage contributions of 3.49% and 3.95%, respectively. Petroleum refining is clearly linked to the oil industry. In addition, producers of government services and other industries rely on oil revenues and government spending. Therefore, the contribution of the private sector to the GDP is arguably overestimated because it is critically dependent on the oil industry. This argument is supported by other empirical evidence (Aldukheil, 2013; Alodadi & Benhin, 2015; Niblock & Malik, 2007; Ramady, 2013). For instance, Alodadi and Benhin (2015) examined the significance of the non-oil industry in contributing to economic growth over the period 1970 to 2011. Dividing the determinants of economic growth into two parts—oil and non-oil exports—they found that oil exports were responsible for economic growth while non-oil exports were not economically significant.

Figure 2.1 shows the percentage contributions of the oil, private, and government sectors to the GDP from 1970 to 2016.

Figure 2.1: Percentage Contributions of Institutional Sectors to Gross Domestic Product from 1970 to 2016 at Constant Prices (2010=100)



Source: Based on Saudi Arabian Monetary Authority (2016) data.

As can be clearly seen in Figure 2.1, the private sector tends to correlate with the oil sector, emphasizing the susceptibility of the private sector to oil price volatility. Therefore, it can be argued that a large part of the Saudi private sector is insignificant and unstable, which could be attributed to the backward non-oil sectors—a typical example of the so-called “Dutch disease.”

In the early phase of Saudi Arabia’s economic development, a “project approach” was adopted where the focus was on specific projects, funded or supported by the government. This is in contrast to a “sectoral approach” where a whole sector is developed, in accordance with specific indications (Niblock & Malik, 2007, p. 59). The only sectors that could have unmistakably undergone sectoral development were the oil and financial industries. The development of the financial sector could be attributed to its sensitivity and to the fact that it is under the direct supervision and responsibility of only one public agency—the Saudi Arabian Monetary Agency (SAMA). It appears that this has been reflected in greater efficiency of the sector, by alleviating coordination issues and conflicts of interest arising from different goals of different public departments, as witnessed in other sectors. Moreover, regulations and a dynamic system were developed to mitigate any difficulties facing the sector. SAMA even has fostered

several mergers among small and medium financial intermediaries. For instance, the Saudi Investment Bank in 1976, Al Rajhi Bank in 1978, and Bank Albilad in 2004. Each one of these banks was the result of a merger of several financial intermediaries, aiming to enhance the banks' profitability and the sector's efficiency and stability. The sector has also experienced direct government investment in creating several financial institutions, when the private sector was seemingly reluctant to do so, as for example, in the case of Riyad Bank in 1957.

In contrast, the other sectors were left to the private sector without a "sectoral development" strategy. However, the private sector, for various reasons, fell short and was unable to take the initiative in this regard. Consequently, many non-oil sectors are ridden with hundreds of thousands of fragmented micro stores. As mentioned earlier, these are virtually completely dependent on cheap low-skilled foreign labor and exceedingly low levels of capital expenditure, reflecting poorly on technological diffusion and working conditions. According to the Annual Economic Establishments Survey (General Authority for Statistics, 2018), the relative distributions of small, medium, and big establishments were 83%, 14%, and 3%, respectively. These categories are defined by the number of employees as follows: small (less than 5), medium (5 to 19), and large (20+). To gain a better understanding of the industrial structure of the Saudi economy, the following discussion examines the top four industries and their percentage distributions according to various criteria.

The first-ever census of establishments was the "Enumerating Establishments" census of 2010, issued by General Authority for Statistics (2010a). Based on this census, Appendix 2 contains a comprehensive table representing percentage distributions of industrial groups by different establishment sizes and employment nationality, ranked by industry size. As mentioned earlier, the focus is only on the top four industrial groups according to different criteria. Table 2.2 thus lists the top four industrial groups by percentage distributions of the total number of establishments.

Table 2.2: Top Four Industries by Percentage Distributions of the Total Number of Establishments

Industry	Percentage
Wholesale and retail trade; repair of motor vehicles	47.62%
Manufacturing	10.94%
Accommodation and food service activities	10.55%
Agriculture, forestry and fishing	9.95%

Source: Percentages were calculated by the researcher based on the Census of Establishments (General Authority for Statistics, 2010a)

Table 2.2 indicates that almost 50% of total establishments are from the wholesale, retail trade and repair of motor vehicles industrial group, ranked as the largest industrial group in the Saudi economy by number of establishments. The other top three industrial groups with the highest number of establishments are (i) manufacturing, (ii) accommodation and food service activities, and (iii) agriculture, forestry, and fishing with percentage distributions of 10.94%, 10.55, and 9.95%, respectively. A large gap can be seen in the percentage distribution between the first- and second-ranked industrial groups. It would be ideal to link those industries with their contributions to the GDP in Table 2.1. Unfortunately, the industrial categories are not consistent with each other. For instance, Table 2.1 includes restaurants and hotels with wholesale and retail trade and there is no indication of the repair of motor vehicles. Despite this shortcoming, the wholesale and retail trade industry contributed only 4.63% to the GDP on average over the period 1970 to 2017—even with the inclusion of restaurants and hotels. This might be considered a modest contribution, given that almost 50% of establishments in the Saudi economy are from that industry.

The research indicates that the micro stores model is one primary reason behind the low productivity of the private sector. For instance, Table 2.3 lists the top four industrial groups by percentage distribution of micro establishments.

Table 2.3: Top Four Industries by Percentage Distribution of Micro Establishments

Industry	Percentage
Agriculture, forestry and fishing	94.57%
Other service activities	93.28%
Real estate activities	91.96%
Wholesale and retail trade; repair of motor vehicles	88.45%

Source: Percentages were calculated by the researcher based on the Census of Establishments (General Authority for Statistics, 2010a)

Table 2.3 shows that the wholesale industrial group is ranked fourth, with about 88% of its establishments being micro stores. Micro stores tend to create dead-end jobs because of the limited number of employees and because of the non-complex nature of production, such as shopkeepers in corner grocery stores, mechanics in a motor repair shop, or other maintenance shops employing less than five workers. These are considered casual/traditional businesses, which are unlikely to innovate or create meaningful job opportunities for citizens. The low capital intensity also reflects negatively on technological diffusion, and thereby, low productivity.

Similarly, Table 2.4 and Table 2.5 list the top four industrial groups by percentage distribution of medium and large establishments, respectively.

Table 2.4: Top Four Industries by Percentage Distribution of Medium Establishments

Industry	Percentage
Financial and insurance	48.70%
Education	44.91%
Mining and quarrying	44.53%
Human health and social work activities	44.49%

Source: Percentages were calculated by the researcher based on the Census of Establishments (General Authority for Statistics, 2010a)

Table 2.5: Top Four Industries by Percentage Distribution of Large Establishments

Industry	Percentage
Mining and quarrying	44.74%
Human health and social work activities	31.19%
Education	29.95%
Electricity, gas, steam and air conditioning supply	26.47%

Source: Percentages were calculated by the researcher based on the Census of Establishments (General Authority for Statistics, 2010a)

As can be seen in Tables 2.4 and 2.5, apart from mining and quarrying and the financial and insurance industries, all other industries generally contribute insignificantly to the GDP. As mentioned above, the oil and financial sectors have undergone sectoral development that likely contributed to their high efficiency and productivity.

Turning to employment distribution, Table 2.6 demonstrates the distribution of Saudi versus non-Saudi workers in the top four industrial groups.

Table 2.6: Percentage Distribution of Saudi Versus Non-Saudi Workers in the Top Four Industries in Terms of Total Establishments

Industry	Industry Percentage	Saudi Percentage	Non-Saudi Percentage
Wholesale and retail trade; repair of motor vehicles	47.62%	20.73%	79.27%
Manufacturing	10.94%	20.08%	79.92%
Accommodation and food service activities	10.55%	14.86%	85.14%
Agriculture, forestry and fishing	9.95%	14.48%	85.52%

Source: Percentages were calculated by the researcher based on the Census of Establishments (General Authority for Statistics, 2010a)

As demonstrated in Table 2.6, evidently, the top industrial groups by number of establishments are important employers of non-Saudi workers. The low costs of foreign labor combined with the low start-up costs of micro stores reveal that the private sector has adopted a low capital-intensive model, which in turn reflects

on the low productivity of the sector and incapability of generating meaningful job opportunities for the relatively skilled Saudi labor force. In addition, there is no difference in the share of Saudi employees relative to the important changes in the establishment distributions of the first and second industries. This may contest the argument that the retail industry can play a significant role in recruiting Saudi labor. There must be something other than the number of establishments by which the growth of Saudi employment in the private sector is induced.

In contrast, Table 2.7 shows that about 80% of total establishments are concentrated in four industries.

Table 2.7: Top Four Industries and the Percentage Distribution of Workers of the Same Nationality

Industry	Industrial Percentage Distribution of Total Establishments	Saudi Percentage Distribution of Total Saudi Workers	Non-Saudi Percentage Distribution of Total Non-Saudi Workers
Wholesale and retail trade; repair of motor vehicles	47.62%	24.05%	27.71%
Manufacturing	10.94%	13.91%	16.68%
Accommodation and food service activities	10.55%	5.34%	9.21%
Agriculture, forestry and fishing	9.95%	3.55%	6.32%
Total	**79.05%**	**46.85%**	**59.92%**

Source: Percentages were calculated by the researcher based on the Census of Establishments (General Authority for Statistics, 2010a)

As Table 2.7 reveals, these four industrial groups are responsible for recruiting about 47% and 60% of the total Saudi and foreign workers, respectively. However, it is important to bear in mind that Saudis are reported in the Census of Establishments as employees while they are most likely owners who engage in rent-seeking activities. In other words, many Saudi people create low-risk

small businesses and rent out those businesses to guest workers for fixed monthly payments, either as an additional income besides their regular jobs or as the main source of income. Subleasing stores to foreign workers holding work visas, and not investor visas, is illegal and known as “commercial concealment” (discussed in section 3.2.3). Regardless of this shortcoming, the analysis of the Saudi market continues assuming Saudis as employees, while keeping in mind that the Saudi employment rates in the private sector are most likely to be overestimated.

The top four industrial groups with the most national workers along with industrial and labor percentage distributions are presented in Table 2.8.

Table 2.8: Top Four Industries with the Most National Workers Along with Industrial and Labor Percentage Distributions

Industry	Industrial Percentage Distributions of Total Establishments	Labor Share of Total Workers in All Industries	Saudi Labor Share of Total Workers in an Industry
Electricity, gas, steam and air conditioning supply	0.07%	1.05%	75.51%
Financial and insurance	0.56%	1.55%	68.28%
Information and communication	0.49%	1.42%	67.21%
Mining and quarrying	0.06%	1.92%	65.27%

Source: Percentages were calculated by the researcher based on the Census of Establishments (General Authority for Statistics, 2010a)

As indicated in Table 2.8, electricity, gas, steam and air conditioning supply is the most localized industry (measured by the number of Saudi workers), with 75.51% of its total employees being Saudis. The other three most localized industries are (i) financial and insurance, (ii) information and communication, and (iii) mining and quarrying, with the percentage of Saudi workers at 68.28%, 67.21, and 65.27%, respectively. What is striking about the employment structure of the Saudi industry is that the least localized (i.e., least Saudized) industries are the least concentrated ones, and vice versa, the most localized

(i.e., most Saudized) industries are the most concentrated ones. This can be attributed to various reasons. One is that companies in these localized industries were mostly public institutions before privatization. As public institutions, they were important employers of Saudi workers. Despite becoming privatized, the government continues to own a large share in some cases. For example, according to the Saudi Stock Exchange (2019), the Public Investment Fund owns 74.3% of the Saudi Electricity Company and 70% of the Saudi Telecom Company. Therefore, government ownership may play an important role in employment decisions.

Another likely reason is that if we look carefully at each one of these four industries, we will notice that the government has granted concessions to some companies and/or embarked on initiatives through joint ventures with the private sector to develop a certain industry. For instance, the Saudi market has only one oil company, one gas company, one electricity company and 11 banks, and it had only one telecommunication company before the government granted a concession to three more companies. Moreover, the Public Investment Fund owns 70% of the Saudi Basic Industries Corporation (SABIC), 10.91% of the National Gas and Industrialization Company, 44.29% of the National Commercial Bank, and so on (Saudi Stock Exchange, 2019). These companies are large and tend to be producers and distributors of goods and services; they also adopt advanced technology, allowing them to operate efficiently and recruit relatively skilled workers. Hence, they may enjoy many advantages that are correlated with large firms, notably, greater efficiency and profitability and higher compensation. This is consistent with the empirical findings of Krueger and Summers (1986), Schmidt and Zimmermann (1991) and Shaffer (2009). For instance, Krueger and Summers (1986) state that “high wages tend to be paid in industries that are concentrated, have high profits, and have relatively small labor shares” (p. 26). Manning (2011) concurs, stating that “[i]t is a well-documented empirical fact … that big establishments pay higher wages than small establishments” (p. 1017).

In short, for any reason—either government ownership influencing employment decisions or the government granting a concession to few large

companies, or both— these companies have succeeded to a greater extent in developing the industries in which they operate by diffusing advanced technology, diversifying government revenues, and localizing the labor force. Therefore, they have undoubtedly played an imperative role in economic development, while the regulatory bodies supervise the market and prevent any monopolistic behavior. In many aspects, this is a win-win situation for all stakeholders—employers, employees, the state, and society.

The top four industrial groups with the least national workers, along with industrial and labor percentage distributions, are presented in Table 2.9.

Table 2.9: Top Four Industries with the Least National Workers Along with Industrial and Labor Percentage Distributions

Industry	Industrial Percentage Distributions of Total Establishments	Labor Share of Total Workers in All Industries	Saudi Labor Share of Total Workers in an Industry
Construction	3.28%	16.34%	11.47%
Agriculture, forestry and fishing	9.95%	5.68%	14.48%
Accommodation and food service activities	10.55%	8.31%	14.86%
Arts, entertainment and recreation	0.23%	0.50%	16.03%

Source: Percentages were calculated by the researcher based on the Census of Establishments (General Authority for Statistics, 2010a)

As Table 2.9 demonstrates, the share of Saudi workers in these industries is between 11% and 16% of the total workers in the industry. In general, these industries tend to be dominated by fragmented establishments such as micro stores. For instance, micro stores constitute 54% of establishments in the construction industry, 95% in the agriculture group industry, 77% in the accommodation group industry, and 65% in the arts group industry (see Appendix 2). In addition, except for the fourth industrial group, which is a neglected industry anyway, the other industries listed in Table 2.9 are less concentrated than those

in Table 2.8, which suggests that Saudi employment localization is associated with industrial concentration.

In terms of establishments, Table 2.10 shows the number of establishments and their percentage distributions according to different establishment sizes.

Table 2.10: Distribution of Establishments by Size

	Establishment Sizes by Number of Employees			Total
	Micro (1–4)	**Small (5–19)**	**Large (20+)**	
Total Establishments	677,390	108,017	20,970	806,377
Percentage Distributions	84.00%	13.40%	2.60%	100.00%
Cumulative Number	677,390	785,407	785,407	
Cumulative Percentage Distributions	84.00%	97.40%	97.40%	

Source: Percentages were calculated by the researcher based on the Census of Establishments (General Authority for Statistics, 2010a)

As can be seen in Table 2.10, the percentage distributions of micro, small and big establishments are 84%, 13.4%, and 2.6%, respectively. The table also shows that the cumulative percentage distribution for micro and small establishments is 97.4%. Despite the category of big establishment being an open category and starting from a relatively small number of employees, the share of this group is only less than 3%. This is an alarming percentage, given the low economic performance of the private sector. This reveals the extent to which the Saudi private sector is critically lacking formal large firms that could contribute to “sectoral development.” It is dominated by fragmented informal establishments that are unlikely to have the means to contribute significantly to employment growth, let alone economic diversification or technological progress.

Similarly, Table 2.11 demonstrates the percentage distributions of Saudi, non-Saudi, and total workers in all industries according to different establishment sizes.

Table 2.11: Distribution of Employees by Nationality and Establishment Size

Labor Groups	Establishment Sizes by Number of Employees			Total
	Micro (1–4)	Small (5–19)	Large (20+)	
Saudi Workers	269,407	192,690	572,895	1,034,992
Saudi Labor Share	20.32%	22.04%	25.24%	23.15%
Non-Saudi Workers	1,056,577	681,445	1,697,029	3,435,051
Non-Saudi Labor Share	79.68%	77.96%	74.76%	76.85%
Total Employment	1,325,984	874,135	2,269,924	4,470,043
Percentage Distribution of Total Employees	29.66%	19.56%	50.78%	100.00%

Source: Percentages were calculated by the researcher based on the Census of Establishments (General Authority for Statistics, 2010a)

As can be observed in Table 2.11, the distribution of Saudi versus non-Saudi employees of the total employees is 23.15% and 76.85%, respectively. The Saudi economy is not only dependent on foreign workers but also susceptible to risk exposure to shortages of foreign labor for any reason, particularly those industries dependent on a single nationality and with low substitution elasticity. The risk exposure to shortages of foreign labor is not an impossible event; such a scenario occurred in the past as a spillover effect of the Iraqi invasion of Kuwait in the early 1990s. Hence, the risk exposure to the shortage of a certain nationality may occur in the future too. Here, the risk is being referred to as a coercive or political event; however, shortages may also result from restrictions on foreigners' work visas, in efforts to increase national employment in the private sector. Evidently, such a business environment heightens uncertainty and negatively affects the incentives to invest in the private sector.

Table 2.11 indicates that there is little variation in percentage distributions across different sizes of establishments for both Saudi and non-Saudi employees. Thus, establishment size alone might not be informative as regards Saudi employment. In particular, the percentages relating to Saudi workers may be because of the localization requirements, and not the result of the establishment size per se. Moreover, many Saudis are reported as employees when in fact, they are the owners of establishments, as mentioned earlier.

Further, some 50% of total employees work in micro and small establishments while the remaining 50% work in big establishments. In other words, even though big establishments constitute only 3% of total establishments, they contribute to approximately 50% of total employment, 25% of which are Saudi workers.

In terms of the wage structure in the Saudi labor market, Table 2.12 shows the average monthly wage per paid employee in different sectors by gender and nationality. As can be seen in Table 2.12, there is an average wage differential of more than SR6,200 (approximately USD1,600)[7] between Saudi and non-Saudi employees. When considering the average across different sectors in these two groups, we can see that the differences come from the “private establishments” and “regional and international organizations.” This highlights the gap in the reservation wage between Saudi and non-Saudi workers in the private sector. However, the average wage levels in the “private establishments” may be masked by the wage levels of formal or large firms because they are usually required to report to official institutions, which is not the case for informal businesses. Small businesses in Saudi Arabia do not usually have financial records (Al Hajjar & Presley, 1996). Thus, the average wage levels in “private establishments” are likely to be overestimated.

[7] The Saudi Riyal is pegged to the USD at a fixed exchange rate, SR 3.75 = $1.

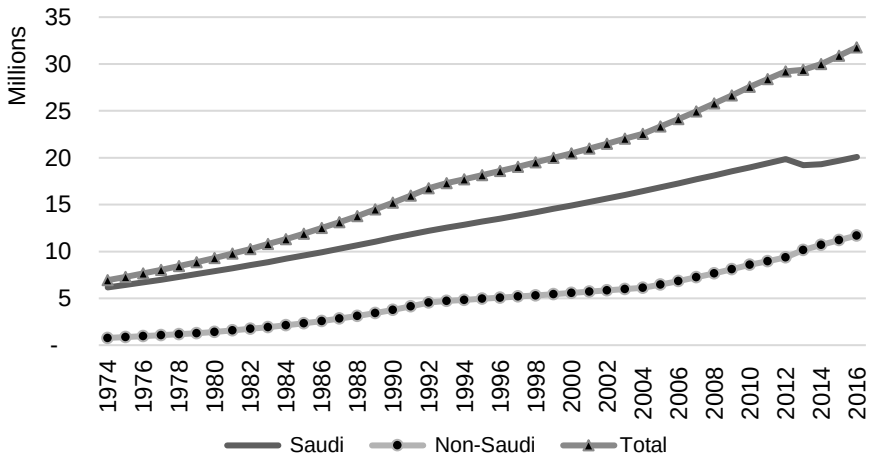
Table 2.12: Average Monthly Wages Per Paid Employee (15+) by Gender, Nationality and Type of Sector in Saudi Riyals

	Saudi			Non-Saudi			Total		
Sector	**Male**	**Female**	**Total**	**Male**	**Female**	**Total**	**Male**	**Female**	**Total**
Governmental	10,921	9,981	10,735	8,913	8,321	8,827	10,765	9,890	10,596
Private Establishments	7,988	5,426	7,570	3,783	5,876	3,812	4,377	5,564	4,423
Non-Profit Organizations	4,811	3,829	4,563	3,592	2,575	3,465	4,017	3,276	3,887
Domestic Labor	0	0	0	2,122	1,577	1,882	2,122	1,577	1,882
Regional and International Organizations	10,739	15,400	12,535	6,612	0	6,612	7,146	15,400	7,766
Total	10,160	8,995	9,939	3,792	2,503	3,674	6,080	6,177	6,093

Source: Labor Market (General Authority for Statistics, 2017b)

Turning to demographic aspects, Figure 2.2 shows both the aggregate population growth and disaggregated population growth by nationality from 1974 to 2016.

Figure 2.2: Total Population by Nationality Between 1974 and 2016

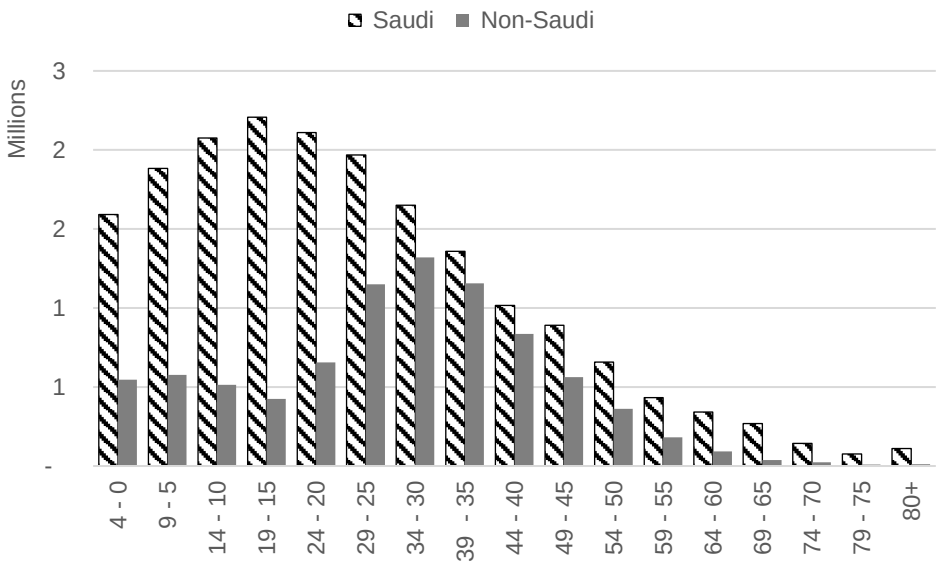


Source: Based on Saudi Arabian Monetary Authority (2016) data.

The Saudi and non-Saudi population grew between 1974 and 2016 at annual growth rates of 2.85% and 6.85%, respectively. Of the total Saudi population, the non-Saudi population constituted about 37% in 2016 compared to only 11% in 1974. The rising number of foreign workers could be attributed to the increasing number of micro stores that create a disguised labor shortage, combined with the sponsorship system facilitating the influx of foreign workers (discussed further in Chapter 3). It is crucial to note the difference between the two types of population growth, because some authors—for example, Spiess (2010)—attribute local unemployment to the high growth rate of the total (aggregated) population. This may be not true because, as mentioned earlier, between 1988 and 2016, the average annual growth rate of new shops was 7% compared to the 2% growth of the indigenous Saudi population (Ministry of Municipal and Rural Affairs, 2018; Saudi Arabian Monetary Authority, 2016). This fact might be obvious but has not been sufficiently emphasized or appreciated in the existing literature.

Figure 2.3 presents the population distribution by age group and nationality.

Figure 2.3: Population by Age Group and Nationality in 2010



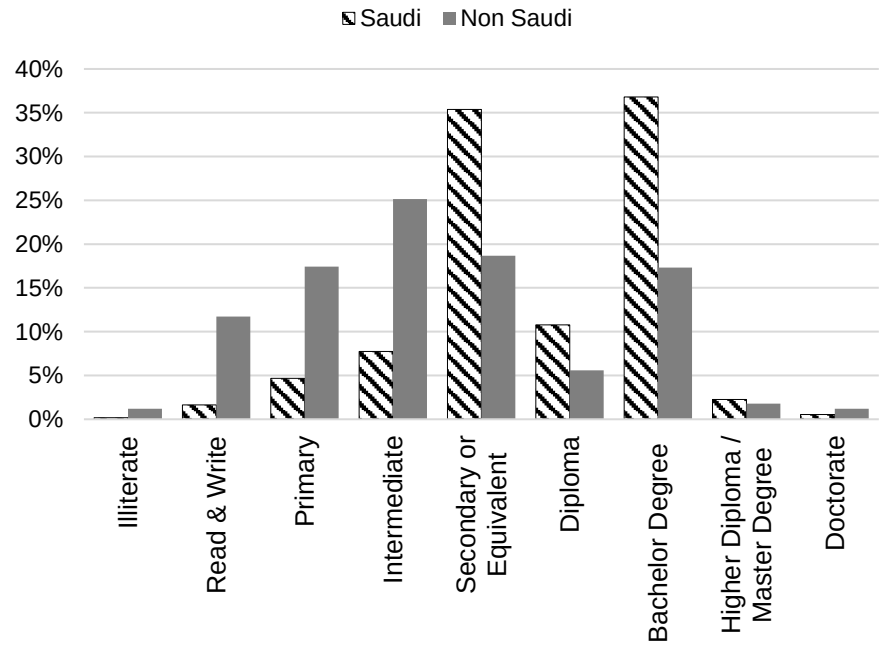
Source: Based on General Authority for Statistics (2010b) data.

As depicted in Figure 2.3, most members of the non-Saudi population are in their thirties, which could be due to the fact that many guest workers reside in Saudi Arabia without their families. In contrast, 63% of the Saudi population is under 30 years old, making it a relatively young population. This young population is sometimes referred to as the “economic gift” or the “window of opportunity”—presumably because it reduces the dependence rate (Chaaban, 2009). Unfortunately, the Saudi economy may be unable to fully exploit this “economic gift” because of the high unemployment among the young population. According to General Authority for Statistics (2019b), of the total local workforce that is unemployed (approximately 12%), 74% and 54% were in the 15–29 and 20–29 age groups, respectively. Given the issue of high unemployment among the youth, the issue of the aging of the population (Abusaaq, 2015), and the issue of early retirement of Saudi workers, Saudi pension funds may face financial difficulties in the future, or even a financial crisis, if this pressing issue is not resolved (International Monetary Fund, 2018). Ramady (2013) pointed out that the Gulf Cooperation Council (GCC) countries, including Saudi Arabia, exhibit demographic imbalances because of the large portion of young people of the

total population or the so-called “youth bulge” and the high number of guest workers. This situation may give rise to several challenges, such as creating jobs for new entrants to the labor market, increasing unemployment, security issues or depressing average wage levels.

Further, for several decades, Saudi Arabia has experienced persistent and high unemployment among its local population, particularly among educated young Saudis. The phenomenon of educated unemployment in Saudi Arabia is supported by different evidence (Ramady, 2010; Stevens, 1986). Figure 2.4 shows the percentage distributions of the labor force by education status and nationality.

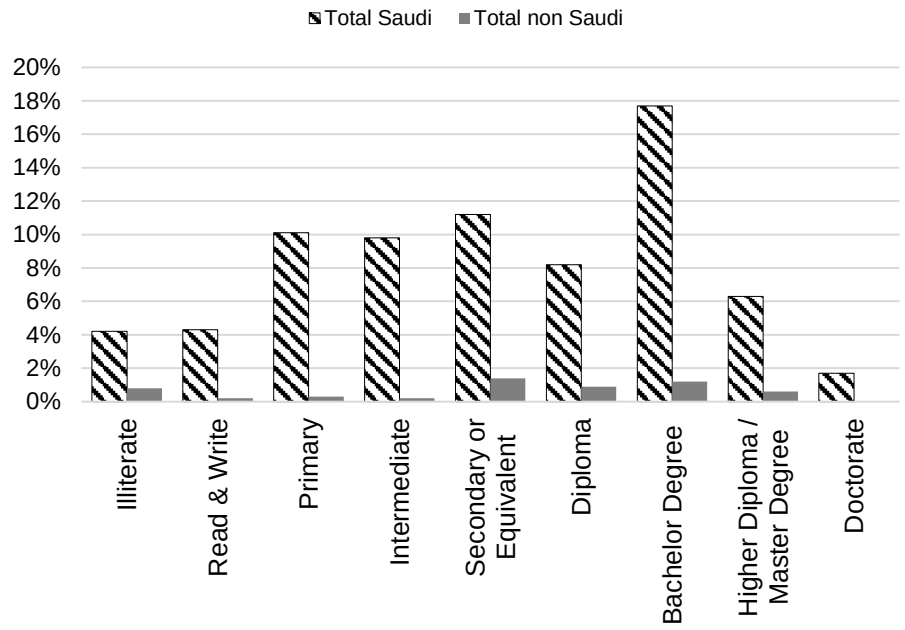
Figure 2.4: Percentage Distribution of Labor Force by Education Status and Nationality



Source: Based on General Authority for Statistics (2017d) data.

As can be seen in Figure 2.4, unlike non-Saudi labor, the Saudi labor force has more years of education, with the majority holding secondary and bachelor degrees. Figure 2.5 presents the unemployment rate by nationality and education level.

Figure 2.5: Total Unemployment Rate (15+) by Nationality and Education Level, Q4 2017



Source: Based on General Authority for Statistics (2017e) data.

Figure 2.5 indicates that the largest percentage of unemployed Saudis is constituted by those who have a bachelor degree. Thus, unemployment in the Saudi workforce is higher among skilled workers. This finding challenges the proponents of the skill mismatch hypothesis, who attribute Saudis' unemployment to their lack of skills. However, as Figure 2.5 reveals, the Saudi labor force is more skilled than the foreign labor force that dominates the private sector. In other words, if the skill mismatch theory had explained unemployment among the Saudi workforce, the unemployment rate would have been higher among the unskilled labor group as opposed to the skilled group. The high unemployment rate among Saudis is a major challenge that has persisted for decades, particularly among young, educated people. This challenge not only prevents the economy from exploiting the "economic gift" but also places it under increasing pressure from other unemployment costs.

Various authors have touched on the different aspects of unemployment costs, cautioning that these may lead to serious consequences. Unemployment costs can be categorized into (i) social costs including social exclusion,

psychological problems, family disruptions, increased poverty and crime (Al-Thaqafi, 2000; Assidmi & Wolgamuth, 2017; Chaaban, 2009; Niblock & Malik, 2007); (ii) political costs (Sadi & Henderson, 2010; Spiess, 2010); and (iii) economic costs. The latter can occur at the microeconomic level—for example, the loss of individuals' labor income and increasing firm costs resulting from localization and expatriate levies that were introduced because of the increasing pressure of unemployment (Peck, 2017). Economic costs can also occur at the macroeconomic level—for example, the loss of national income through a reduction in GDP growth associated with increasing unemployment (Alrasheedy, 2019; Ramady, 2010).

For the economic costs of unemployment, economists use Okun's law pertaining to the negative relationship between the unemployment rate and GDP. The law states that when the unemployment rate decreases by 1%, aggregate output increases by 3% (Oi, 1983). For instance, Ramady (2010) used Okun's law to estimate Saudi Arabia's potential output loss associated with unemployment for the period of 1993 to 2008, and found that the total GDP losses were about SR1,021 billion. Alrasheedy (2019) estimated the loss of real GDP resulting from unemployment using both Okun's law and the average product method (i.e., the real GDP per worker multiplied by the total number of unemployed people), and found that the average product method could be more accurate in the case of Saudi Arabia. Thus, using the average product method, he estimated that the total loss of Saudi output in 2016 was USD95 billion (approximately SR356.25 billion) as a result of 1,687,313 unemployed Saudis (13.7% of the total real GDP) (Alrasheedy, 2019). Regardless of other unemployment costs—which are by no means less important—this is a considerable loss of income, resulting from the persistently high unemployment of Saudis. This state of affairs needs close attention from researchers as well as urgent action from policymakers.

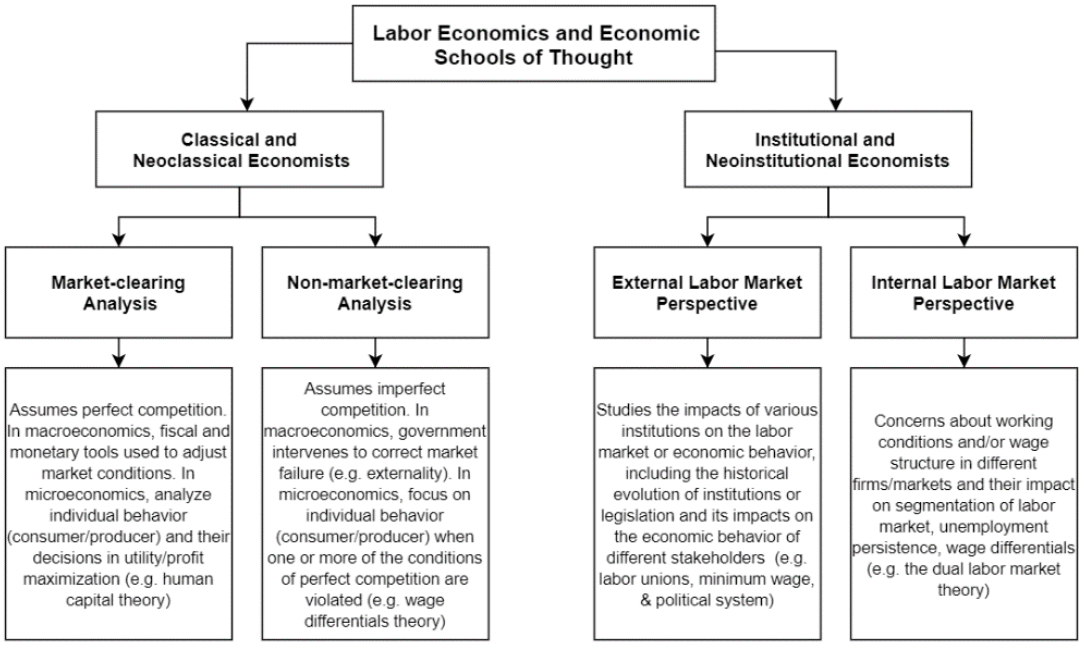
2.3 Economic Schools of Thought and Unemployment

The theory of labor economics has long been a cornerstone in economics literature because of its impacts on both the macro and micro levels of the economy. Indeed, because of the widespread repercussions of unemployment on households, firms, and governments, it is crucial to understand the causal relationship between the unemployment rate and different factors to effectively reduce it. Labor economists usually categorize unemployment into three main types: frictional, structural and cyclical. Ehrenberg and Smith (2012) attribute these different types of unemployment to different reasons as follows:

> Frictional unemployment arises because labor markets are dynamic and information flows are imperfect; structural unemployment arises because of long-lasting imbalances in demand and supply. *Demand-deficient unemployment* is associated with fluctuations in business activity (the "business cycle"), and it occurs when a decline in aggregate demand in the output market causes the *aggregate* demand for labor to decline in the face of downward inflexibility in real wages. (p. 514)

Although there may be general consensus about these three types of unemployment, economists view unemployment from different perspectives, depending on the economic school of thought to which they subscribe. In the literature on labor economics, unemployment is generally addressed through two different schools of thought—classical and institutional economists. Figure 2.6 represents a broad overview of these two schools of thought.

Figure 2.6: Labor Economics and Economic Schools of Thought



Source: Compiled by the researcher

As shown in Figure 2.6, the two schools of thought can be divided into classical and neoclassical economists on the one hand, and institutional and neoinstitutional economists on the other. While the former tend to adopt the positivist or deductive method, embracing rational choice, marginal productivity, and considering the market structure as homogeneous and "given," the latter tend to adopt the normative or inductive method, advocating heterogeneity of the market structure as an essential assumption of inquiry (Boyer & Smith, 2001; Farkas & England, 1988). Further, Figure 2.6 illustrates different analyses and perspectives adopted within each group. Classical and neoclassical economists may subscribe to economic analyses depending on whether they assume a self-clearing market. In contrast, institutional and neoinstitutional economists emphasize the role of institutions in labor markets but they approach the analysis of institutions from either an external (pre-market forces) or internal (in-market forces) perspective.

To elaborate, for market-clearing analyses (competitive models), economists generally assume perfect competition, including complete information; hence, markets would clear and be in a state of equilibrium through the mechanism of

demand and supply. For instance, microeconomic theory focuses on individuals (e.g., consumers/workers and producers/firms) as in human capital theory and leisure-work analyses. Using utility (profit) maximization conditions, they conclude that wages (prices) are determined by the marginal product of labor (capital). This approach excludes other endogenous factors (e.g., tastes) and institutional frameworks from the analysis (Taubman & Wachter, 1986). According to the perfect competition theory, the “law of one price” should prevail across different markets. That is, if a price of a certain good is under the equilibrium level, it would be “heated up” by increasing demand until it reaches the equilibrium level. Similarly, if the price is above the equilibrium level, demand would decline, causing the price to fall to the equilibrium level. The basic premise of competitive model analyses is that prices and wages are flexible; hence, they adjust in various market conditions, leading to equilibrium (Boyer & Smith, 2001; Dequech, 2007; Kerr, 1950; Smith & Zoega, 2009). However, wage, as a price for labor, tends to be inflexible or “rigid,” particularly downward (Samuelson & Nordhaus, 2010). The wage rigidity above the equilibrium level would result in labor supply being higher than labor demand, causing unemployment. Labor market failure to clear has led many economists to believe that the market-clearing model can no longer underpin empirical studies of the labor market (Bhaskar et al., 2002; Card & Krueger, 2017). Economists do not agree upon the causes of wage rigidity.

As for non-market-clearing analyses (non-competitive models), economists acknowledge market disequilibrium and propose theories to explain wage rigidity. However, there appears to be no consensus as to the reasons behind such an economic phenomenon. One explanation for wage rigidity is the efficiency wage hypothesis. Because of imperfect information and transaction costs, it might be considered efficient for employers in certain markets to pay higher wages for various reasons, including an exertion of ability (Marshall, 2013) and the reduction of recruitment and monitoring costs, while treating “equilibrium unemployment as a worker discipline device” (Shapiro & Stiglitz, 1984, p. 433). Another explanation for wage rigidity is the fact that labor markets tend to be segmented. At the same time, there exist some factors preventing labor mobility

from one labor market to another (e.g., from a lower-wage market to a higher-wage one). Because labor market segmentation is related to this research, the rest of this section is devoted to discussing different economic theories attempting to answer a question of why the labor market segmentation does exist.

Classical and neoclassical economists tend to use “non-competing labor groups” instead of labor market segmentation, and then explain different factors contributing to labor immobility between labor markets including occupational, geographical, and industry-specific factors that perpetuate labor market segmentation or wage differentials. For instance, Solow (1980) attributes persistent disequilibrium to wage rigidity resulting from market segmentation, which in turn may be caused by a variety of factors such as transportation, information, and transaction costs. But more importantly, he argues, wage rigidity is caused by the economic rent resulting from the buildup of firm-specific or industry-specific human capital and mutual awareness among market participants of their expectation; hence, the persistence of market segmentation. Another reason for the labor market segmentation is imperfect information and transaction costs. For instance, according to Bulow and Summers (1986), because jobs are difficult to monitor in the primary, but not in the secondary, labor market, wages in the primary labor market tend to be above the market-clearing level while wages in the secondary labor market tend to be at the market-clearing level. Hence, high wages in the primary labor market, as a “dismissal threat” mechanism, would lead to an excess supply of workers for primary jobs, causing segmented labor markets. One implication of Bulow and Summers’s model is that the allocation of workers across industries may be inefficient, resulting in too few primary jobs (Rebitzer, 1993).

In contrast, institutional and neoinstitutional economists may address the non-clearing market or unemployment by adopting either an external (pre-market forces) or internal (in-market forces) perspective of the labor market. The former refers to labor market institutions, such as labor-related regulations, labor unions, and collective bargaining, that affect employment relationships in the labor

market while the latter refers to the internal labor market, as in the dual labor market theory (Loveridge & Mok, 1979).

From an external labor market perspective, the existence of labor market institutions could be attributed to various reasons, including "imperfect information, uneven market power (between employers and workers), discrimination, and inadequacies of the market to provide insurance for employment-related risks" (Betcherman, 2013, p. 1) Some labor institutional factors could be considered as obstacles to clearing markets. For instance, Solow (1980) acknowledges the failure of markets to clear and admits that there may be some non-economic[8] factors contributing to involuntary unemployment. He discusses Pigou's works, *Lapses from Full Employment* and *Theory of Unemployment*, mentioning four institutional factors as obstacles to the classical functioning of the labor market that result in involuntary unemployment. These factors include segmented markets, trade unionism, unemployment insurance, and public opinion and its ensuing effects on legislations. He concludes by advocating the segmented labor market, making valuable remarks worth quoting at length:

> The sort of labor market I have in mind is segmented. It often makes sense to think of an employer or definable group of employers as facing its own labor pool. Some members of the labor pool may be unemployed, but still belong to it. Although transportation, information, and transaction costs are possible sources of segmentation, they need not be among the most important. The buildup of firm-specific or industry-specific human capital may be more fundamental, and equally a kind of mutual knowing-what-to-expect that gives both parties the labor market a stake, a rent, in the durability of the relationship. This point is close to the distinction between auction markets and customer markets made by Arthur Okun in a different context. The labor market, at least the "primary" labor market, is a customer market; this may be one of the

[8] Acknowledging non-economic factors from a prominent neoclassical economist may be an important and surprising remark at the same time. Even Solow (1980, p. 3) himself stated that "the second general point I want to make is one that I am surprised to hear myself making. While I find several of the candidate hypotheses entirely believable, I am inclined to emphasize some that might be described as noneconomic."

> important facts that differentiates the primary from the secondary labor market. (Solow, 1980, p. 9)

McDonald and Solow (1985) identify three distinct sorts of unemployment. First, workers in the primary labor market may be laid off when product demand is depressed. But, for various reasons, they may not move on to the secondary labor market to seek jobs. Second, because of the slow adjustment of wages in the secondary labor market, various levels of frictional unemployment may result. Third, the authors indicate that:

> there is—even in "equilibrium"—a group of workers who have abandoned the secondary labor market and are "waiting" or "searching" or just trying to gain admission to the primary labor market. In the meantime, they are unemployed. It is pointless to dismiss this as "voluntary" unemployment. Our claim is that it is an unintended by-product of the operation of the labor market, not merely a reflection of individual preferences. (McDonald & Solow, 1985, p. 1117)

McDonald and Solow (1985) assert that this last type of unemployment should not be dismissed as "voluntary" because it does not result from "individual preferences." This is a particularly important remark in the context of this study. Someone might consider unemployed Saudis, who are unwilling to participate in the secondary labor market and waiting for a job in the primary labor market, as being "voluntarily" unemployed, while in fact, it is an unintended by-product of the operation of the Saudi labor market dominated by micro stores.

In contrast, proponents of labor institutions argue that "the evidence dispels fears that government or union interventions invariably impede labor-market adjustments." (Freeman, 1993, p. 404). In a later paper, Freeman (2005) contends that there are two reasons for inconclusive debate over the claim of the negative effects of labor institutions on aggregate performance:

> The first reason is that many adherents to the claim hold strong priors that labour markets operate nearly perfectly in the absence of institutions and let their priors dictate their modelling choices and interpretation of empirical results. The second reason is that the cross-country aggregate data at issue is weak,

> too weak to decisively reject strong prior views or to convince those with weaker priors. (Freeman, 2005, p. 130)

From an internal labor market perspective, the focus is on firm-specific administrative factors creating the so-called dual labor market. The dual labor market theory was first developed by Doeringer and Piore (1970) when explaining the issues of low wage levels and unemployment. They argue that the labor market can be divided into two parts: the primary and secondary labor markets in which jobs and workers have different characteristics. In the primary labor market, jobs may be considered “good jobs” because they usually have positive characteristics such as high wages; hence, they offer substantial returns on human capital, stable/secure employment, good working conditions, internal labor market (i.e., rule-based management within a firm), opportunities for promotion, and equity. In the secondary labor market, jobs may be considered “bad jobs” because they usually have negative characteristics such as low wages, poor working conditions, high labor turnover, few opportunities for promotion, and arbitrary management. Unlike secondary jobs, primary jobs are “rationed,” meaning not all workers who are qualified for primary jobs can find one. This implies that the advantages of those rationed jobs will increase prolonged unemployment among people who have the financial ability to wait or think they deserve a primary job. In other words:

> even if the secondary market is perfectly competitive, there will be a wage differential between the primary and secondary market, and some unemployment will persist at equilibrium. (Demekas, 1990, p. 850)

Moreover, workers’ characteristics can be distinguished in these markets according to their skills, tendency for turnover, lateness, and absenteeism. Therefore, several characteristics can be used to distinguish between these two labor markets, such as wages and fringe benefits, job security, opportunities for training and promotion, regular working hours, labor turnover, skill and qualification requirements, firm size, and predetermined administrative rules within establishments. These characteristics are better or more abundant in the

primary labor market than in the secondary labor market (Dickens & Lang, 1988; Doeringer & Piore, 1970; Harrison & Sum, 1979).

In short, wage rigidity contributing to unemployment is due to labor market segmentation, while certain factors prevent labor mobility from one labor market to another. Labor immobility could be attributed to three factors: occupational, geographical, and skill specialization factors (Ehrenberg & Smith, 2012; Farkas & England, 1988). These factors prevent or reduce movement between markets, changing the balance of labor demand and supply between the markets, and ultimately, causing excessive labor in some markets (i.e., unemployment) and shortages in others.

It is worth highlighting the following points before concluding this section. First, a clear distinction between different economic schools of thought is difficult since ideas, theories, and methodologies are evolved and shared between different schools. Second, even within an economic school of thought, theories may not be agreed upon by all economists who subscribe to the same school. Third, “orthodox versus heterodox” terminology could have been used to describe the different schools of economic thought. However, such an approach would have been even more controversial and beyond the scope of this research. Moreover, it has been investigated thoroughly by other authors (see, for example, Colander et al., 2004; Dequech, 2007). The next section discusses the Saudi labor market, particularly with regard to unemployment issues.

2.4 Discussion of Prior Studies on Saudi Unemployment

Prior studies on Saudi unemployment generally adopt three theories or models: the human capital theory, the institutional theory, and socioeconomic theories. The following three sub-sections are devoted to each one of these theories.

2.4.1 Human Capital Theory

The human capital theory postulates that the most important factors determining an individual's wage are formal education, on-the-job training, and work experience, which are assumed to be significant and positively related to the individual's productivity (Baffoe-Bonnie, 2003). Experience refers to years of working, including internship and apprenticeship programs. Researchers tend to distinguish between general training and specific training as two separate forms of human capital investment. The first pertains to general education while the second pertains to on-the-job training. These two forms are discussed in the following sub-sections.

2.4.1.1 General Training

Human capital theory can be explained as the study of "activities that influence future monetary and psychic income [i.e., consumption] by increasing the resources in people. These activities are called investment in human capital" (Becker, 1993, p. 11). Education and on-the-job training are two important forms of investment in human capital because they increase knowledge and skills, thereby increasing earnings, productivity and reducing the probability of being unemployed (Becker, 1993). Some authors attribute increasing wages to "credentialism" rather than to schooling per se. While Becker (1993) acknowledges the existence of credentialism, he argues that it does not explain most of the positive association between earnings and schooling, as suggested by the evidence. Regardless of the causality issue between education and the level of wages and (un)employment, the positive effects of human capital investment on employment and wage levels seem to be established in the literature. But the question is whether we can attribute the high Saudi unemployment rate to insufficient skills of Saudi labor. The human capital theory has been used in many previous studies investigating the issue of Saudi unemployment. Hence, numerous researchers attribute chronic high unemployment to insufficient skills among Saudis. It may be true to some extent

that insufficient skills play a part in this issue; however, considering it as the ultimate cause is questionable.

One issue related to Saudi human capital is whether Saudi labor has the "sufficient skills" required by the labor market. Several studies suggest that Saudi workers do not have the skills required by the private sector (Al-Asfour & Khan, 2014; Baqadir et al., 2011; Harry, 2007; International Monetary Fund, 1997). For instance, Baqadir et al. (2011) used questionnaires and semi-structured interviews to investigate whether the recent changes, in the 2000s, in vocational training addressed the perceived skill gap in the Saudi labor force. They grouped the perceived skills into three categories: work ethics, specialized knowledge, and generic skills. They found that a skills gap may indeed exist. Further, the common perception of private sector employers is that technical education fails to offer Saudi students sufficient vocational training to a level that employers require or expect. However, the following question could be posed: How could foreign workers be more skillful given that they generally lack a proper education or are even illiterate, and have language barriers? Undeniably, many Saudi workers do lack important skills, but they are still relatively more skilled than foreign labor, as discussed in section 2.2. To put things into the right perspective, we need to bear in mind the following issues.

First, it seems there are some exaggerations or fallacies among employers in the private sector. Specialized knowledge can be gained with experience, but Saudi labor faces difficulty in finding jobs in the first place because they are likely not given an equal opportunity for work. In contrast, other categories of skills can be acquired in a short time if employers are willing to recruit Saudi workers. Moreover, work ethics and generic skills should be already familiar to any high-school graduate, and most of the Saudi labor tends to have a college degree or vocational training (more evaluation of work ethics under socioeconomic theory).

Second, it seems there is a misperception, particularly among Saudi employers, that college graduates should be suitable for any task without any

orientation or initial training at the beginning of recruitment (discussed further in the on-the-job training section).

Third, it is essential to bear in mind the possible bias in some studies that have used surveys to measure skills among Saudi workers. This is because economists usually measure skills by years of schooling and then use this as one variable among others in a multivariant regression to estimate the returns of education on wages in different industries or groups of people (Angrist & Krueger, 1999). However, in the case of Saudi Arabia, previous studies tend to ask employers a direct question as to whether Saudi workers have the required skills. This might lead to measurement error since the definition of skills may differ between employers and researchers. Besides, descriptive analysis (sees Figure 2.4 and Figure 2.5) shows that the level of years of schooling is higher among the Saudi labor force compared to non-Saudis and the unemployment rate is higher among Saudi workers with higher years of schooling, which may cast doubt on the argument of insufficient skills of Saudi labor. Another reason for possible bias in these studies is that employers may state anything to circumvent the regulations regarding guest workers' visas and localization requirements. Anecdotal evidence suggests that employers tend to use insufficient skills to circumvent the regulations regarding guest workers' visas and localization requirements. If this were not the case, how is it possible that they claim insufficient skills in Saudis while at the same time recruiting unskilled foreign workers?

Finally, 97% of establishments in the Saudi market are regarded as micro and small establishments (General Authority for Statistics, 2017a). These establishments create jobs that do not normally require specialized or sophisticated skills and are considered low-paid and lacking in security and promotion opportunities. In effect, Saudi labor may arguably be considered overqualified (specialized) compared to most jobs created in the private sector because of the predominance of micro and small establishments. It could be argued that the Saudi market has undergone a structural shift, resulting in changed conditions of labor demand and supply, through the emergence of the

low capital-intensity business model that depends on cheap and low-skilled labor. That is, jobs created in the private sector, particularly by micro establishments, do not match the skills that Saudi labor has acquired throughout their years of education. Hence, in Chapter 3, the study offers an alternative theoretical framework by considering the structure of the Saudi labor market and market-distorting factors that might have contributed to the labor market duality and perpetuated unemployment among the national labor force.

Another issue related to Saudi human capital is whether the quality of the education system is sound and whether it is relevant to labor market requirements. Because of the importance of human capital investment in various development aspects, the Saudi government has spent generous amounts on the education system over the last half century.[9] However, several researchers question the quality and relevance of the education system. Hence, they call for improving syllabi, providing labs, and fostering innovative thinking and problem-solving skills rather than teaching subjects that are unrelated to labor market requirements, such as Islamic studies or humanities (Mishrif & Alabduljabbar, 2018; Niblock & Malik, 2007; Sadi & Henderson, 2010). This is a valid concern, especially regarding improving school buildings and students' skills as well as integrating the English language, computer skills and programming into the syllabi. Nevertheless, there have been attempts to downplay the importance of Islamic studies, which could be criticized for two reasons. First, because of the Saudi institutional structure, graduates of Islamic studies play an imperative role in fulfilling thousands of jobs in various institutions in both the public and private sectors, such as schools, juridical institutions, or consultancy and law firms. This might be the only comparative advantage of Saudi labor, allowing them to compete with non-Saudi labor—not only in the public sector but also in the private sector, given the status quo of the Saudi market structure. This important fact has not been appreciated in the existing literature.

[9] For example, the Aleqtisadiah newspaper reported on January 8, 2019 that the Saudi government had spent about SR1.2 trillion (approx. USD320 million) on education in 2004–2013 (Abdullah et al., 2013).

Second, it is not clear how other researchers arrived at the conclusion that there are too many Islamic studies graduates because, as part of the literature review of this study, no empirical studies could be found supporting this claim. Some recent statistics, such as those presented in Table 2.13, only support the claim regarding humanity studies (i.e., Literary).

Table 2.13: Percentage Distribution of Saudi Unemployed Persons (15+) Holders of Secondary Education or Equivalent by Sex and Educational Specialization

Educational Specialization	**Male**	**Female**	**Total**
Science	61.3	55.7	58.3
Literary	35.1	43.8	39.8
Industrial / Professional / Area	1.9	0.0	0.8
Health	0.0	0.0	0.0
Commercial	0.9	0.0	0.4
Religious sciences	0.9	0.5	0.7
Agricultural and technical	0.0	0.0	0.0
Total	100	100	100

Source: Labor Market (General Authority for Statistics, 2019c) The percentage distribution of unemployed Saudis with secondary education or equivalent is about 28.4%. There is seemingly no similar data for bachelor degree holders, who constitute 56.4% of all unemployed individuals.

Table 2.13 represents the percentage distribution of unemployed Saudis who hold secondary education or an equivalent qualification. It can be seen that unemployment among science graduates is higher than that of religious/Islamic sciences. Assuming these statistics accurately reflect the relevance of educational specializations to the labor market requirements, such a distribution of unemployment could be due to two key reasons. First, the private sector lacks economic diversity and job creation because science graduates rank first among unemployed groups. To some extent, this is supported by news reports. For instance, several reports indicate increasing unemployment among dentistry graduates (Hazzazi, 2019), engineers (Al Zahrani, 2017), and even graduates from overseas universities (Al-Maliki, 2015). The latter would certainly not have gone abroad to specialize in Islamic studies. Second, the public sector is the prime employer of Saudi labor and the fact is, specialization in Islamic studies is

the main requirement for many jobs in the public sector. Both explanations are possible because the private sector lacks the diversity that accommodates such specialties while the public sector is, indeed, the primary employer of Saudi labor. Therefore, there is no wonder that educational institutions have continued offering such programs. We cannot blame educational institutions for offering programs required by most jobs and institutions nor can we blame the youth for pursuing the degrees in higher demand. After all, this is what human capital theory suggests, that education and on-the-job training, as forms of human capital investment, should be directed to the highest expected returns—which Islamic studies happen to be in the case of Saudi Arabia.

The lower unemployment of certain educational specializations could indicate the relevance of that specializations to labor market requirements. However, these statistics may not be instructive because they may not necessarily reflect the relevance to labor market requirements nor the unemployment period. There are four conceivable cases where such statistics would not necessarily reflect the relevance of a major to labor market requirements. First, the number of graduates with a certain major could be greater than the number of available jobs. Yet statistics still show lower unemployment because of the possible existence of underemployment issues, as is the case with many jobs in the public sector. In this case, we would mistakenly conclude the relevance of that majors whereas in actual fact, we should reevaluate the provision of it. Second, the number of graduates in a certain major could be less than the number of related jobs, while statistics still show higher unemployment because of the lack of diversified industries that could service different specializations or because employers favor foreigners over Saudi workers in the private sector. In this case, we would mistakenly conclude the irrelevance of that major whereas in fact, there is a need for such a major; however, the problem lies elsewhere. Third, the number of graduates of a certain major could be greater than the number of related jobs. But while statistics show higher unemployment, the unemployment period might be shorter than others—possibly because of a short lag between job and graduate growth rates. In this case, we might mistakenly conclude the irrelevance of that educational major whereas in fact, graduates play a crucial

role in fulfilling opening jobs. Fourth, the number of graduates of a certain major could be less than the number of related jobs. However, while statistics show lower unemployment, the unemployment period might be longer—possibly because of an extended lag period in the related job growth. In this case, we might mistakenly conclude the relevance of that major. Therefore, because of labor market distortion, those descriptive statistics might be not instructive when it comes to the relevance of educational specializations to labor market requirements.

A final issue discussed here and related to human capital theory in the Saudi context is whether Saudi Arabia's human capital investment emphasizes higher education at the expense of technicians and mechanics. In contrast to proponents of increased investment in human capital through higher education and overseas scholarship programs as the solution to Saudi unemployment, Stevens (1986) points out that the government's decision to increase human capital investment might be considered a "very rapid decision" (p. 22) that had not been carefully thought through. He maintains that there were "too many PhDs and too few plumbers," resulting in educated unemployment and disguised unemployment in the public sector (Stevens, 1986, pp. 21–22). Nowadays, the situation may be still the same. In other words, there is still a shortage of technicians, while many Saudi job seekers who studied abroad face difficulties finding a job upon returning to Saudi Arabia even though they trained in countries such as the USA, the UK or Australia. Hence, the issues of educated unemployment and underemployment on the one hand, and a shortage of technicians on the other, have become more prominent than ever. However, it can be difficult to develop specialized and professional workers while the overall market size is small and undiversified. This is because market size may constrain investment in skills, as explained in the next sub-section. It is even more difficult with the low capital-intensive business model that prevails in the private sector. Unless there is a demand by professional firms offering well-paid jobs for trained technicians, the Saudi youth will not have enough incentives to choose a vocational education path.

2.4.1.2 Specific Training

In addition to education as a form of human capital investment, there is so-called on-the-job training that plays a crucial role in improving specific skills and transferring expertise. However, it is barely mentioned in previous studies addressing the lack of skills in the Saudi labor force. Prior research seems to emphasize educational reform as a sure way of solving the problem of persistent unemployment among citizens. Evidently, many researchers and employers have high expectations of the education system, to the extent that they expect graduates to be “instant specialized workers” without any need for further (specific) training. However, Gary Becker, the Nobel Prize laureate and father of human capital theory, asserts that “[e]ven college graduates are not well prepared for the labor market when leave school, and they are fitted into their jobs through formal and informal training programs.” (Becker, 1993, p. 20). Al-Dosary (1991) draws attention to a major myth pertaining to the formal education system, viewed as responsible for specific training. He states that the consensus among specialists is that the formal education system is only responsible for training students to be ready for on-the-job training, by which they can obtain further practical training, and then become workers with specific skills.

The private sector similarly tends to neglect on-the-job training. This may be due to several reasons. First, employers may be unaware of the importance of training to increase productivity. Second, employers may be worried about workers resigning before they offset the cost of their training. Third, the private sector seems to be relying on government-funded programs for training job seekers and new entrants into the labor market. But the issue of ineffective training programs can be raised here too. The first two issues can be mitigated by raising awareness and improving contracts while the third can be mitigated through a re-evaluation process after obtaining feedback from employees and employers. However, this research argues that the most important reason for the absence of a training culture in the private sector is because this sector is small, undiversified, and dominated by micro stores. These three facts were established in section 2.2 as the main characteristics of the private sector, but

they deserve a little more elaboration in relation to human capital investment and specific training.

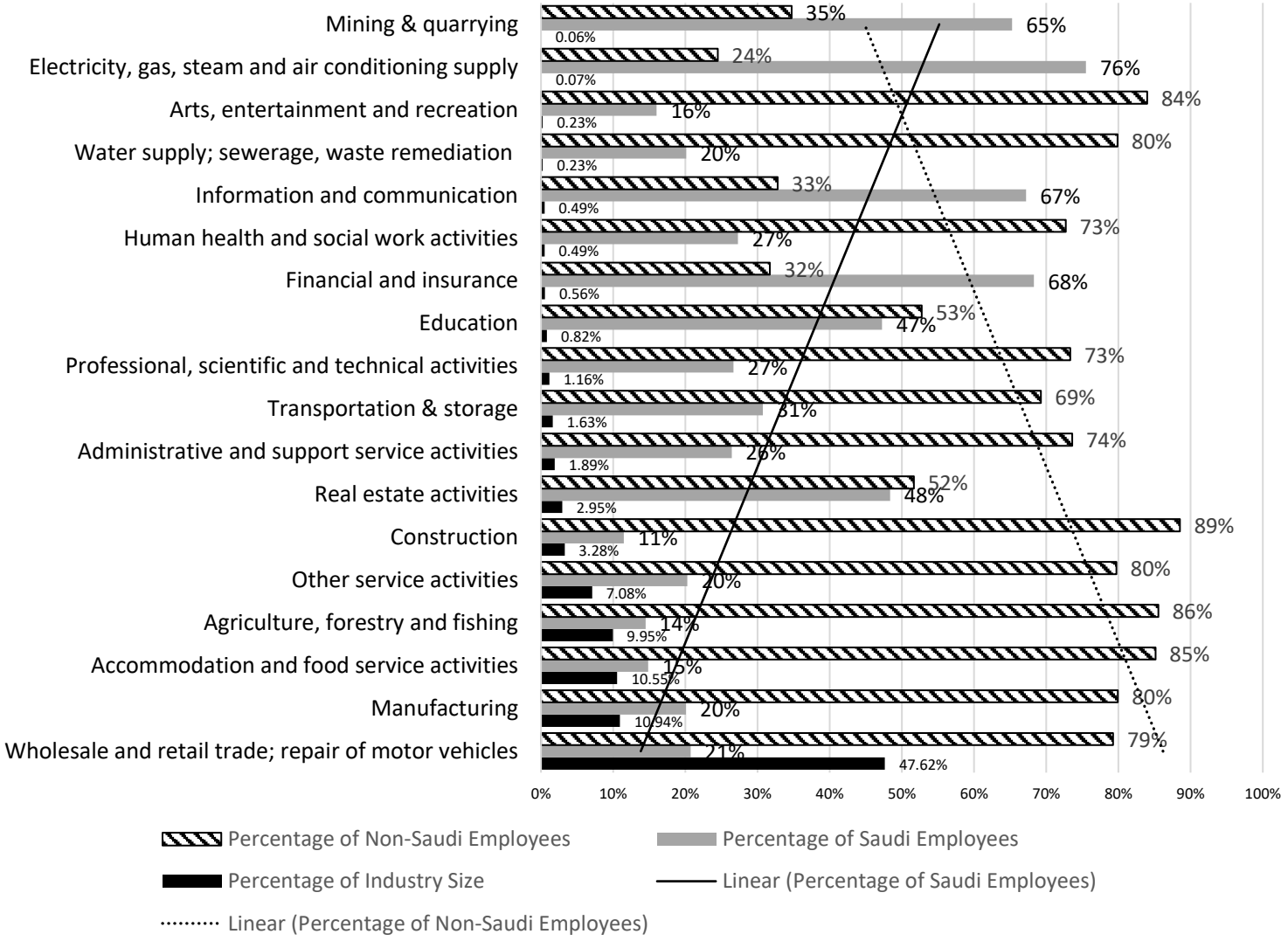
Compelling evidence suggests that formal and informal training (e.g., casual training by co-workers) tends to occur in large firms since they can utilize scale economies, thereby reducing training costs (Black et al., 1999). Moreover, Becker (1993) highlights the influence of market size on specialization and the incentive to invest in skills. He maintains that “[s]pecialization in an activity would be discouraged if the market were very limited; thus the incentive to specialize and to invest in oneself would increase as the extent of the market increased” (Becker, 1993, p. 88). In addition, Haber et al. (1988) investigated employment and training opportunities in small and large firms and concluded that “[b]ecause of the importance of monitoring costs and of efficiencies that result from the routinization of production when producing large standardized volumes of output, large firms tend to provide firm-specific training” (p. 89). Finally, Oi and Idson (1999) reviewed the labor economics literature on firm size and wages; citing Barron et al. (1987), they state that ”[t]he chances of receiving any of five different kinds of training (formal, informal, by coworkers or managers) were higher in larger establishments and in multi-plant firms … The available international evidence goes in the same direction” (p. 2204).

Because it is dominated by micro and small establishments, the Saudi private sector has exceedingly low levels of technological investment. For instance, 97% of total establishments fall into the category of micro and small businesses (Table 2.10). Unlike large firms, production in micro and small establishments is not based on advanced technology requiring specific training. Moreover, the descriptive statistics reveal that the median value of the net capital formation is about SR645,000 only (see Table 5.1), which highlights the extent of low capital investment in the private sector. If we subscribe to the argument that the current education system does not provide the “right” or “required” skills, why do we not see on-the-job training? This would, at least partly, offset this deficiency and build up specific training, as occurs in large firms such as ARAMCO or SABIC. This is because recruiting and training Saudi labor is costly, while micro and small

establishments do not generally require specific training. And in case where they do require some sort of training, they do not have economies of scale and scope to reduce the average costs of recruiting and training Saudi labor.

To elaborate on how the Saudi undiversified market may negatively affect employment and human capital investment, Figure 2.7 ranks different industries in the private sector from the most to the least concentrated. First, Figure 2.7 shows that the private sector is largely dominated by the so-called “Wholesale, retail trade, repair of motor vehicles” industry. This industry is considered the least concentrated because 48% of total establishments in the Saudi market are in that group. The next three least concentrated industries (Manufacturing, Accommodation, and Agriculture) range between 7% to 11% of total establishments. Second, it can be noticed, as mentioned in section 2.2, that the least localized (i.e., least Saudized) industries are the least concentrated, and vice versa, the most localized (i.e., most Saudized) industries are the most concentrated. In fact, there is a negative linear relationship between the number of Saudi employees and the number of establishments (the solid trend line). By contrast, there is a positive linear relationship between the number of non-Saudi employees and the number of establishments (the dashed trend line). This can be attributed to the fact that jobs in the least concentrated industries tend to be created by large, professional firms that are able to recruit and train Saudi labor, as mentioned in the discussion of Table 2.8. Third, those least concentrated industries (e.g., Wholesale) tend to be dominated by micro stores, require no specific skills, and recruit unskilled cheap foreign labor. By contrast, the most concentrated industries (e.g., Mining) tend to be dominated by large firms, require specific skills, and recruit skilled Saudi labor. This fact alone could provide enough evidence to refute the hypothesis of insufficient skills of Saudi labor.

Figure 2.7: Industrial Diversification and Labor Share by Nationality in the Saudi Private Sector



Source: Based on the Census of Establishments (General Authority for Statistics, 2010a). For more details, see Appendix 2.

Such conditions of the Saudi market have been reflected not only in low investment in general and specific human capital development but also in the reduced effectiveness of the implementation of internship and apprenticeship programs. For instance, Spiess (2010) describes the Saudi labor market, stating that:

> the concept that young people learn life skills and gain valuable experience through volunteer work, extracurricular activities and summer employment is almost completely absent in the region. (p. 15)

Therefore, the fact that the private sector is small, undiversified, and dominated by micro stores contributes largely to the issue of low Saudi employment as well as the low human capital investment in specific training through on-the-job training programs in the private sector.

In short, the Saudi labor market seems to lack a collective effort between policymakers, educational institutions, and employers. We should not emphasize human capital while overlooking other important market forces. This does not mean that we should not always endeavor to enhance the education system and the skills of Saudi labor. However, emphasizing human capital development without enhancing the market size and creating opportunities for “meaningful” jobs, may exacerbate unemployment, particularly educated unemployment, and underemployment.

2.4.2 Institutional Theory

While the role of the labor market is to foster mutually beneficial transactions enhancing economic efficiency, the government’s role is to make certain transactions mandatory according to certain values praised by the society (Ehrenberg & Smith, 2012). Labor market institutions may be viewed through two different lenses: “institutionalism” or “distortionism.” The former holds that

“institutions can reduce transaction costs, enhance productivity, and moderate crises” (Betcherman, 2013, p. 2), and often criticizes empirical analysis by challenging its assumptions. The latter measures the impact of labor market institutions without considering related historical and social factors, and views institutions as distorting labor markets, thereby impeding economic efficiency (Betcherman, 2013).

Examining the impacts of labor market institutions is crucial because of its potential distributional effects. Thus, although labor market institutions are meant to regulate the market and mitigate market failure, they could have unintended consequences that impede economic efficiency. Labor market institutions should be arranged or devised so that benefits outweigh costs. This is easier said than done; there has been a decades-long debate about the optimal institutional framework. The following sub-sections review three aspects of labor institutions related to the Saudi labor market: Saudi labor law, localization policies, and Saudi labor market segmentation.

2.4.2.1 Saudi Labor Law

According to Al-Rayes and Al-Abed (2017), the *Worker Compensation System in Industrial and Technical Projects* dating back to 1937 was the first kernel of the Saudi labor system. Despite its limited applications, it was required, at the time, to regulate the earliest influx of foreign workers, particularly those who worked in oil drilling and exploration. However, the labor system issued in 1942 was the first comprehensive Saudi labor system, covering all workers and more labor-related issues. Since then, the system has undergone several issuances or amendments, dated as follows: 1947, 1969, 2005, and lastly, 2015 (Al-Rayes & Al-Abed, 2017).

Several studies have investigated the impact of Saudi labor law on the business environment. For instance, some authors focus on the importance of

increasing transparency, reducing bureaucracy and fostering an investment-friendly environment (Al-Dosary & Rahman, 2005; Zamberi Ahmad, 2012), embracing an entrepreneurship model (Kayed & Kabir Hassan, 2011), comparing labor regulations between the private and public sector (Al-Buraik, 1988; Mahdi, 2000), or efficiency of localization policies (Al-Dosary & Rahman, 2005; Al-Sudani & Abdulkheer, 2001; Peck, 2017; Ramady, 2013).

One aspect of previous studies on the Saudi labor market is whether there is a legal distinction between the different sectors or workers. According to Mahdi (2000), labor regulations differ between the private and public sector. He claims that while most labor regulations apply only to the public sector, the statement of personal/casual contract usually governs the relationships between employers and employees in the private sector. Such contracts are not uniform, reflecting labor regulations. They may vary—not only among different firms but also among employees in the same firm. However, these issues could be attributed to the informal business relationship between employers and employees in the private sector because of the prevailing micro stores model. After all, in case of any dispute between labor parties, they must uphold the Saudi labor law.

Another aspect of prior studies examining the Saudi labor market is whether dismissing Saudi workers is more difficult than non-Saudis. Since this is a controversial issue and is frequently raised by researchers, highlighting the inflexibility of Saudi recruitment, it might be useful to quote some articles from Saudi labor law. First, Article (2) mentions different terms, including “worker” which is defined as “[a]ny natural person, whether male or female, working for an employer and under his management or supervision for a wage, even if he is not under his direct control” (Ministry of Labor, 2019, p. 18). Second, according to the latest amendment by Royal Decree No. (M/64), dated 5/6/1436 H. corresponding to (25/3/2015), Articles (75–77) state the following:

> Article (75): If the contract is of an indefinite term, either party may terminate it for a legitimate reason to be specified in a written notice served to the other party at least thirty days prior to the termination date if the worker is paid monthly

> and not less than fifteen days for others. Article (76): If the party terminating the contract does not observe the period provided for in Article (75) of this Law, such party shall pay the other party compensation equal to the worker's wage for the duration of the notice, unless both parties agree on a larger amount. Article (77): Unless the contract contains specific compensation against termination by either party for an illegitimate reason, the aggrieved party is entitled to compensation as follows: 1. A 50-day pay for the worker's each year of service, if the contract is not of a fixed term. 2. The wage for the remaining period of the contract if the contract is of a fixed term. 3. Compensation referred to in paragraphs (1) and (2) of this Article may not be less than two months' wage. (Ministry of Labor, 2019, pp. 46-47)

It is evident that these articles do not discriminate between Saudi and non-Saudi workers nor do they suggest inflexibility of labor recruitment. On the contrary, it could be argued such articles may not reserve workers' rights (Saudis and non-Saudis alike), especially since they tend to be the marginalized party. According to (Al-Rayes & Al-Abed, 2017), the latest amendment to Article 77 abrogates the authority of the labor judiciary to estimate the compensation. They consider this approach as arbitrary because it is not based on logical or legal justifications nor is it equivalent to the damages to the worker resulting from the contract termination. Therefore, the claim of inflexible labor recruitment may not be quite true. In fact, it may emphasize the idea of job instability, rather than inflexibility, in the private sector.

In short, it could be argued that apart from localization policies, there is no explicit discrimination in the Saudi labor law between workers in the two sectors or between Saudi and non-Saudi labor. However, two main features could be considered discriminatory between Saudi and non-Saudi labor, namely, wages and employment eligibility. The latter stems from localization policies, while the former may be attributed to minimum wage regulations and wage structure. This may be more relevant to the public sector than the private sector, which has no unifying wage structure. Moreover, there are several advantageous features of the public sector, stemming from the structure of public institutions rather than Saudi labor law itself. Adopting employment policies that emphasize preferential employment of Saudi citizens in the public sector may be the prime reason for Saudi workers' preference for the public sector over the private sector (discussed further in section 3.2.1). This does not mean that we should abolish the benefits

of the public sector, but rather, we should call for enhanced working conditions and improved labor regulations in the private sector. For instance, legislation that regulates internship, apprenticeship, and on-the-job programs and training would be of great importance for Saudi human capital development and transferring expertise to the youth and new entrants to the labor market. Further, legislation could regulate certain industries to improve professionalism and work standardization through, for example, professional licenses, job titles and descriptions, wage structures or wage expectations to improve transparency, and uniform contracts reflecting the Saudi labor law while maintaining flexibility. Such efforts would enhance labor rights and reduce antagonistic labor relationships. However, as the study argues, this is unlikely to be achieved given the informality of the micro stores model.

2.4.2.2 Localization (Saudization) Policies

It is important to clarify the meaning of localization in the context of the Saudi labor market since it might have different meanings in different literature, as in industrial organizations theory. In the context of the Saudi labor market, localization policies are simply the regulations that enforce firms to replace foreign workers with Saudi workers according to percentages or quotas. Among other goals, the policy aims to reduce over-dependence on foreign labor as well as the high unemployment rate among Saudis (Sadi & Henderson, 2005). This policy is well known as the “Saudization policy,” and was first introduced in the 1970s. However, it was not enforced until about the mid-1990s (Al-Asfour & Khan, 2014; International Monetary Fund, 1997; Sadi & Henderson, 2005). The Saudization policy has been replaced by a quota system (*Nitaqat*) since May 2011. The following two sub-sections discuss the failure of the Saudization policy, and then, the *Nitaqat* system.

Saudization Failure

The government has been trying to enforce the Saudization policy for more than three decades. Some researchers argue that the policy has failed because of the following reasons. First, the costs of Saudi labor are higher than non-Saudi labor (Ramady, 2010; Sadi & Henderson, 2010). However, the argument of higher costs is challenged by other researchers. For instance, although foreign labor is paid lower wages, there are additional expenses involved in recruiting foreign labor, for example, visa issuance, administrative costs, and termination or transferring costs. Therefore, it could be argued that foreign labor costs more than Saudi labor (Al-Thaqafi, 2000; Kapiszewski, 2000). This argument may be more convincing after the recent expatriate levy introduced in 2017. However, it is also important to consider the difference between various job positions. For instance, in higher-end jobs, non-Saudi workers usually earn higher wages than Saudis because they presumably have more experience or specific skills. In middle jobs, by contrast, non-Saudi workers usually earn less than Saudi workers. Lastly, lower-end jobs usually pay much lower wages, so much so that such jobs do not attract Saudi labor. Given the level of education of Saudi labor, it could be argued that middle jobs are most likely to match their skills, but they find it too difficult to compete with non-Saudi labor. Another even more acute problem is that those different job levels tend to exist in relatively large professional firms; however, such firms are scarce in the Saudi market. According to the size distribution of firms, only 3% of firms are considered large firms that can offer different levels of jobs that require various skills and specializations. In contrast, the remaining 97% of firms are micro or small firms, offering lower-end, low-paying jobs, which are unappealing for Saudi labor.

Second, pressure from powerful businessmen, whose personal wealth was built on cheap foreign labor, are an important factor in delaying the implementation, and later, the failure of the Saudization policy (Ramady, 2013). Third, there is the attitude of the Saudi labor force toward certain jobs which are avoided, mainly by women, for social or religious reasons (Ramady, 2010). Although Looney (1991) agrees that social attitudes play an important role in Saudi labor participation decisions, he maintains that it does not fully explain the

low participation in the private sector (further discussed in section 2.4.3). Again, this study argues that the dominant micro stores model with poor working conditions makes most jobs in the private sector inferior or undesirable. This gives rise to several issues, including low participation and employment of Saudi labor in the private sector, particularly by women.

Fourth, from the point of view of private employers, dismissal regulation is harsh, to the extent that it deters them from recruiting Saudi workers in the first place (Al-Ghaith & Al-Maashoug, 1996; Ramady, 2010). However, as discussed in the previous section, there might some exaggeration in this issue. Finally, Saudi labor is considered, especially by employers, as lacking specific skills and work ethics (Baqadir et al., 2011; Sadi & Henderson, 2005, 2010). Regardless of what employers mean by "work ethics," this could be an outcome of much deeper problems pertaining to working conditions and/or discrimination against Saudi employees—either by Saudi employers or by foreign workers. Discrimination of foreign managers against Saudi employees has also been mentioned by several authors (Al-Dosary, 2004; Al-Ghaith & Al-Maashoug, 1996; Al-Thaqafi, 2000).

Quota System (*Nitaqat*)

Replacing Saudization by the *Nitaqat* system can be considered as an implicit confession of the failure of the Saudization system by the Ministry of Labor. The main difference between Saudization and *Nitaqat* is that in the Saudization system, business activities were classified into only 11 industrial activities and required at least 30% of employees to be Saudi workers regardless of establishment size and industry. In contrast, the *Nitaqat* system first classifies business activities into 41 industrial activities. Then, establishments in each industry are classified into five sizes. *Nitaqat* quotas depend on industrial classification and establishment size (Ramady, 2013).

Introducing the *Nitaqat* system has prompted several researchers to ask whether this new system has a better chance of success. In other words, if we

assume there is a willingness to work by Saudi labor, and to recruit by employers, is there enough Saudi labor supply to offset foreign labor supply? And do the benefits outweigh the costs? Localization policies may have succeeded in the public sector for several reasons, including the non-profit-seeking nature of the public sector as well as the adoption of Saudi recruitment as a policy to mitigate the prolonged unemployment in the country (Kapiszewski, 2000; Ramady, 2010). However, if localization policies were successful in the private sector, it would only be for the short term, and not the long term (International Monetary Fund, 1997; Peck, 2017). In the long term, several compelling pieces of evidence demonstrate that localization policies can lead to counterproductive effects on the economy (Al-Asfour & Khan, 2014; Al-Dosary, 1991; Al-Juhani, 1996; Al Akayleh, 2016; International Monetary Fund, 1997; Peck, 2017; Ramady, 2010, 2013). Localization policies can lead to counterproductive effects because of several reasons, as follows:

First, localization policies may undermine the competitiveness of domestic firms at regional and global level (Ramady, 2013; Sadi & Henderson, 2010). Such policies have negative effects on businesses by increasing operational costs and even forcing some of them to relocate to other countries (Peck, 2017; Ramady, 2013).

Second, localization policies not only increase costs but also increase uncertainty resulting from sudden and reoccurring changes in localization requirements. This increases the probability of capital flight and firm bankruptcy. Peck (2017) contends that the *Nitaqat* system has a significant impact on exit rates. He estimates that the system has caused about 11,000 firms to shut down, raising exit rates from 19% to 28% over the course of about 16 months—from July 2011 to October 2012.

Third, localization policies can create conflicts with other economic policies, such as (i) increasing foreign direct investment, (ii) complying with the World Trade Organization and the International Labor Organization recommendations

to foster market liberalization, and (iii) diversifying and creating a knowledge-based economy as well as encouraging knowledge spillover from interactions with skilled foreign workers (Hertog, 2006; Ramady, 2010).

Fourth, localization policies can slow down economic growth, creating "production bottlenecks" because of limited national labor (Kapiszewski, 2000; Ramady, 2010).

Fifth, localization policies do not differentiate between job positions nor do they take into account whether Saudi workers prefer to take on certain positions for social or other reasons (Ramady, 2013).

Finally, localization policies may compel the private sector to place Saudi workers in marginalized positions, which in turn increases underemployment and undermines productivity. Yet worse, recruiting Saudis simply to comply with localization requirements and not requiring them to come to work results in a practice known as "bogus Saudization" or "ghost workers" (Al-Asfour & Khan, 2014; Assidmi & Wolgamuth, 2017; Hertog, 2018; Odrowąż-Coates, 2015).

All in all, instead of focusing on how to create job opportunities for an increasing labor force, localization policies adopt a "replacement mechanism"—despite experience and empirical evidence proving the ineffectiveness of such policies. Attempting to localize sales jobs at vegetable/produce markets, gold retailers or taxi drivers are good examples of the failure of the "replacement mechanism." Moreover, several authors argue that the elasticity of substitution between Saudi and non-Saudi workers is low, making the implementation of localization policies difficult, if not impossible (Abdalla et al., 2010; Fasano & Goyal, 2004; International Monetary Fund, 1997).

Considering these implications of localization policies, numerous researchers caution that such policies could function as distorting policies, and not as

corrective policies for market failure. Indeed, because local unemployment and Saudi labor turnover rates are still high and the Saudi participation rate in the private sector is low, particularly for women, it can be concluded that localization policies have failed to create *permanent* and decent jobs for citizens in the private sector. Unless we understand the characteristics of various markets that have profound effects on the elasticity of labor supply and demand as well as the elasticity of substitution between different laborers (Saudi vs. non-Saudi), the Saudi economy is likely to continue in a vicious cycle, unable to reduce the high rates of both Saudi unemployment and foreign-labor dependency. Therefore, increasing Saudi employment by creating decent jobs is one of the most pressing issues that policymakers should consider, rather than enforcing localization policies that may create collateral damage for the economy.

2.4.2.3 Labor Market Segmentation in Saudi Arabia

This sub-section examines the labor market segmentation that exists in Saudi Arabia. Labor market segmentation can be addressed from several angles, for example, public versus private sector, nationals versus non-nationals, skilled versus unskilled labor, and male versus female labor. The factors that have contributed to Saudi labor market segmentation are summarized as follows.

First, there is the inter-regional migration of Saudi people (Sherbiny, 1984a; Stevens, 1986). The internal migration of Saudis may have been more apparent during the early economic development of the 1970s and 1980s because modernization policies were concentrated in a few cities. These policies created pull factors in some places and push factors in others, which in turn reduced the geographical and occupational labor mobility of Saudis. Second, there is the adoption of a temporary immigration policy (the guest worker model) (Ahmed, 1993; Fasano & Goyal, 2004; Sherbiny, 1984a). Countries adopting a temporary immigration policy tend to have more pronounced sectoral segregation than countries adopting a permanent immigration policy (Müller, 2003). Third, there is the ability of the public sector to provide good working conditions (Ahmed, 1993;

International Monetary Fund, 1997). Fourth, there is the skills mismatch between Saudi labor and the requirements of the private sector, particularly in upper-end and lower-end jobs (Fasano & Goyal, 2004; International Monetary Fund, 1997).

The idea of stratifying labor skills has also been raised by a number of authors. For instance, Al Sheikh (2001, as cited in Ramady, 2013), suggests that the wide differences in wage levels between Saudi and non-Saudi workers can be explained in terms of two distinct labor markets. He goes on to explain that the supply elasticity of foreign labor varies according to three groups of skills: (i) foreign workers who exhibit a relatively inelastic supply curve, such as specialized and technical workers; (ii) foreign workers who exhibit a more elastic supply curve, such as middle-managerial workers; and (iii) foreign workers who exhibit an almost perfectly elastic supply curve, such as manual workers.

In short, the Saudi labor markets arguably face imbalances of demand and supply of labor, which can be attributed to labor immobility between labor markets. There are three factors contributing to labor immobility, namely, occupational, geographical, and industry-specific skills factors. These factors can create labor shortages in some labor markets, and excessive labor in others. Labor immobility can be attributed to labor market segmentation, which, in turn, stems from various reasons. This thesis argues that three market-distorting factors (discussed thoroughly in Chapter 3) have distorted the Saudi labor market by creating labor market segmentation and imbalances in labor demand and supply, ultimately leading to persistent, high unemployment rates, among other labor issues.

2.4.3 Socioeconomic Theory

This section discusses various factors mentioned in the literature that affect the labor decisions of Saudi individuals (e.g., employers, employees, the unemployed and discouraged workers). These factors are social and economic

in nature; hence, they are discussed within the context of socioeconomic theory. Below is a summary of these factors although there is some overlap in the discussion.

First, there is wealth or non-labor income. Several authors suggest that the high per capita income of the Saudi population and the dependence of the young Saudis on their parents could play a role in their reluctance to accepting entry-level jobs (Birks & Sinclair, 1979; Spiess, 2010; Wilson et al., 2012). Relatively high wealth or non-labor income of individuals on the one hand, and low wage levels in the private sector on the other, means that unemployed people prefer unemployment to working for a wage perceived to be below the subsistence level.

Second, there are low incentives in the private sector. The public sector tends to offer higher wages than the private sector. Moreover, the private sector has inferior job characteristics, such as low wages, low job security, low opportunities for advanced jobs, and unregulated working hours (e.g., more than eight hours a day). Because of sectoral differences, unemployed people may prefer to wait for opening jobs in the higher-wage bracket (Hertog, 2018; International Monetary Fund, 1997; Kapiszewski, 2000; Ramady, 2010), which is likely to have increased unemployment rates.

Third, there is the limited role of capital-intensive projects in employment growth. Looney (1988) claims that capital-intensive projects, such as petrochemical industries, have a limited capacity for jobs, which in turn, has a limited effect on employment growth. This claim may be true in the case of projects undertaken by the government because the private sector has extremely low levels of technological advancement. The private sector has adopted a labor-intensive model (i.e., the micro stores model) that suffers from poor working conditions, including exceedingly low wages and lack of job security and safety; hence, the sector relies almost solely on cheap foreign labor. In other words, without foreign workers—who consider this work environment similar, if not

better, than that in their home country—such a business model may not survive. This may appear as an exaggeration; however, the private sector indeed finds it difficult to survive every time the localization policy is enforced and/or the restrictions on visas of foreign workers become more stringent. In addition, the most localized industries (in terms of national employees) are the most concentrated and capital-intensive (as mentioned earlier when discussing Table 2.8). Hence, before we consider attracting Saudi labor to labor-intensive industries, the government and the private sector may need to improve the work environment and job standards by adopting new technologies and introducing higher wages (this approach is discussed in section 3.4.2). Therefore, even if we agree that capital-intensive investments have limited employment capacity, they at least tend to create valued jobs and have a greater chance of recruiting and training Saudi labor. Moreover, unlike the prevailing labor-intensive industries that depend on foreign labor, capital-intensive industries have a positive impact on Saudi employment growth.

Fourth, the costs of national labor are higher than those for guest workers (International Monetary Fund, 1997). However, as discussed earlier, this view has been challenged by other authors (Al-Thaqafi, 2000; Kapiszewski, 2000). Fifth, there is a lack of effective internship and apprenticeship programs (Baqadir et al., 2011; Spiess, 2010). The absence of such programs is likely to negatively affect the transferal of expertise and practical skills of Saudi labor, consequently lowering their chances of finding jobs in the private sector. Sixth, the rural–urban migration of Saudis or the imbalanced distribution of the population creates geographical immobility of the labor force (Al-Filali & Gallarotti, 2012; Stevens, 1986; Wilson et al., 2012).

Seventh, an important social factor related to labor decisions is that the Saudi labor force has a negative attitude toward working in the private sector. According to Mellahi (2007), the Saudi labor market has four main features, namely, “high population growth, heavy reliance on foreign workers, negative stereotype of local workers, and social perceptions towards work in the private sector” (p. 88). Madhi and Barrientos (2003) state that higher education students

prefer humanities majors to vocational education, and white-collar jobs to blue-collar jobs. Similarly, some authors mention the issue of consumer-oriented society and claim that Saudi youth, in particular, underestimate the value of being employed (Birks & Sinclair, 1979). In principle, Saudi workers may not mind working in the private sector. However, people with a college degree expect more than just working as shopkeepers or other unsophisticated jobs, and they should not be blamed for their ambition. Further, recent evidence suggests that Saudi workers have taken on blue-collar jobs in large firms in the professional industry since they offer good incentives, such as reasonable wages and working hours, and other good working conditions (Ramady, 2014). Moreover, several authors attribute Saudis' negative attitude to the current conditions in the private sector rather than social norms. For instance, some researchers attribute increasing number of foreign workers in the Saudi labor market to issues such as depressed wage levels, which in turn, fosters the negative attitude of Saudi labor toward working in the private sector (Al-Dosary, 2004; Feess, 2012; Ramady, 2013; Sirageldin et al., 1984).

In addition, there are other considerations that could exacerbate this issue, including (i) the high expectations of Saudi workers because of high educational qualifications and the relatively high income per capita, and (ii) Saudi men are typically the breadwinners in the family and are responsible for almost all household expenses. After graduation, young Saudis tend to get married, buy a car, and buy or rent a house. While such items require significant expenditure, young Saudis wish to achieve these milestones within a short time. Hence, when they compare these costs with expected earnings in the private sector, they despair and prefer being unemployed rather than working in the private sector. These two factors, among others, may play an important role in economic behavior with regard to individuals' work decisions.

Along the same lines, some authors maintain that the low participation rates of Saudi women in the workforce is due to Islamic teachings. They argue that the teachings prevent or discourage women from working. However, because there is a significant difference between observance of Islamic restrictions and the

prevention of women's work, Al-Dosary (1991) rejects the notion that Islamic teachings account for poor female participation rates in the labor market, and instead emphasizes the role of "traditional cultural values and norms" (pp. 27–28).

Moreover, many Saudi women work in various professional industries in the private sector in which jobs are considered Islamically compliant. It has been argued that Islamic compliance increases operational costs. This may be true to a certain extent. However, Islamic restrictions are arguably enforced to ensure that women have their space in which to work comfortably, and to prevent or mitigate any type of harassment in the workplace, thereby encouraging their participation in the labor market. Islamic restrictions on women's work are similar to any "safety codes" or "codes of conduct" chosen by different societies or organizations according to their regulations or values. Although these codes increase operational costs, it would be unacceptable to call for abolishing these codes just to avoid incurring these costs. Therefore, Islamic restrictions should also be treated as "codes of conduct" that are needed in the current business environment even though they are likely to increase operational costs. Understanding these costs (i.e., complying with Islamic restrictions of recruiting women) in this way puts things into the right perspective. This assists in finding innovative ways for female workplaces to encourage greater participation of women in the private sector. In particular, the Saudi labor market is dominated by males while many Saudi women avoid working in such workplaces, hence, the labor participation of Saudi women among the lowest in the world. According to Peck (2019), costs relating to women recruitment are considered fixed one-time costs. However, because, among other challenges, many Saudi businesses are small size and have an exceedingly limited capacity of scale economies, they find it difficult to reduce the average costs of recruiting female workers. Peck (2019) proposes several policies programs including "conducting information campaigns to clarify labor regulations, supplying guidelines for upgrading workspaces, facilitating human resources workshops on female hiring and retention, and sponsoring workspace feminization grants" (p. 6). Additionally, this research suggests that these costs can be mitigated by providing "female-only,

shared workplaces", and then renting these workplaces to different establishments. This would encourage participation and recruitment of Saudi women, while meeting the needs of small and financially constrained businesses. As another advantage of this proposal, other supplementary services, such as childcare and transportation, could be provided effectively at lower average costs by such workplaces. Instead of providing grants for each business individually to establish female-only workplaces, the government could fund the establishment of such workplaces and earn profits. The government could also facilitate joint ventures or cooperative funding among businesses to establish such new shared workplaces.

Furthermore, the problem may not be the social or Islamic restrictions per se; there may be other, "confounding" factors that exist, affecting women's decision to work. For instance, these factors could include the lack of so-called "pink-collar jobs" in a male-dominated market as well as the lack of other supplementary services such as childcare, public transportation and the ban on women driving[10]—all of which may have contributed significantly to the low participation of Saudi women in the private sector. Further, Looney (1991) suggests that cultural attitudes are not the main obstacle preventing Saudi women's participation in the labor market. He mentions other factors such as "(a) the country's manpower situation, (b) the need for income at the family and individual level, (c) the encouragement provided by the state, and (d) the skills that job-seekers possess" (Looney, 1991, p. 677).

A final social factor related to labor decisions is that there may be a lack of a strong work ethic among Saudi labor, contributing to the reluctance of private employers to recruit them (Baqadir et al., 2011; Ramady, 2010). However, regardless of what private employers mean by a "work ethic," this claim can be considered as an exaggeration for several reasons.

[10] Women have never been allowed to drive cars in Saudi Arabia until the ban was lifted in June 2018.

One reason pertains to work ethics of foreign workers themselves, who often quit jobs without giving their employers prior notice. In this way, many expatriates become illegal workers by quitting their legal jobs and engaging in various activities in the underground economy, which is a recurrent issue prevalent among foreign workers. However, it would appear that this issue does not deter Saudi employers from recruiting foreign labor, despite being aware of such an eventuality, knowing that they would lose administrative fees and face other expenses that would not apply had they recruited Saudi labor.

Another important reason is the potential conflict of interest between different labor parties—Saudi employers, Saudi labor, and foreign labor. Lindbeck and Snower (2001) mention the “insider versus outsider” notion that has important implications for the labor market. This concept can be applied to the Saudi labor market to explain the tensions between insider employees (foreign labor) and outsider potential employees (Saudi labor). To maximize their benefits, incumbent employees (insiders) try to keep new entrants (outsiders) out of a firm by using tactics such as being unfriendly and uncooperative to the entrant and/or influencing the entrant’s productivity (this is because insiders are usually responsible for training the entrants). For instance, the surveys that substantiate the claims regarding Saudi laborers’ lack of work ethics are usually filled by private employers or by foreign workers, who may not necessarily provide reliable information. Moreover, as mentioned earlier, discrimination against Saudi labor, either by Saudi employers or by foreign workers, has been reported in several studies (Al-Azzaz & Yousef, 1999; Al-Ghaith & Al-Maashoug, 1996; Al-Thaqafi, 2000). In short, this thesis is not attempting to altogether dismiss the idea that some Saudi workers may lack work ethics. However, the issue needs to be approached more objectively, by bringing in important considerations that could assist our analysis of the employment relations in the Saudi labor market.

Overall, although there may be social factors affecting employment and work decisions, previous studies lack an adequate explanation of why such attitudes exist. Pointing to a single factor in a complex array of issues is not helpful and does not provide a better understanding of the labor market. It is crucial to bear

in mind that there are pecuniary as well as non-pecuniary factors that have profound effects on labor decisions. There are also cultural differences between researchers and participants that might distort information or findings. For instance, when researchers who are unfamiliar with Saudi society discuss social issues, they may be prone to some exaggeration. According to Oliver (2010), cultural differences can be a barrier to valid interpretations of the data; hence, findings can be misleading. Therefore, such factors should be considered when addressing job preferences from a behavioral perspective to enhance our understanding of why many Saudis abstain from certain jobs in the private sector.

2.5 Literature Gap

Although previous studies on the Saudi labor market and unemployment have potentially useful implications for the Saudi economy, many suffer from one or more of the following limitations.

First, some studies lack a consistent economic analysis explaining why unemployment persists despite government support for education and businesses to increase Saudi employment in the private sector, and despite the improved skills of Saudi labor in recent decades. Hence, it can be argued that previous studies do not fully explain unemployment persistence. The present study proposes three market-distorting factors that have contributed to a dual labor market within the private sector and perpetuated unemployment among local labor.

Second, many studies focus only on the labor supply theories (e.g., skill mismatch hypothesis) while neglecting labor demand (e.g., market structure and working conditions). Such a limited approach has been proven inadequate given the persistence of high unemployment. The skill mismatch hypothesis only explains a portion of the unemployment, particularly for the higher-end jobs or those requiring specific expertise. However, it does not, as this study argues,

explain unemployment persistence. Focusing only on labor supply theories may exacerbate educated unemployment and prevent us from utilizing the insights derived from other theories. For instance, previous studies fail to apply theories on labor market segmentation, such as the dual labor market theory, which seems to reflect the situation of the Saudi labor market by explaining both the low national labor participation in the private sector as well as unemployment persistence among local labor.

Third, some prior studies lack an estimation of the CES between Saudi labor, capital, and foreign labor in different establishment sizes. The present study uses the nested CES production functions to estimate the elasticities of substitution to evaluate the substitutability between Saudi and non-Saudi labor in different establishment sizes. This approach provides empirical evidence to assess the localization policy as well as understanding the different relationships between different inputs of production.

Finally, many studies have ignored the effect of micro stores on (un)employment, despite the prevalence of this phenomenon in the Saudi labor market. Prior research has also failed to address unemployment in the Saudi labor force from the industrial organization perspective despite the dramatic changes to the market structure over the past decades.

It seems inconceivable to address the issues characterizing the Saudi labor market (e.g., high chronic unemployment, low rates of Saudi employment and participation in the private sector, scarcity of meaningful jobs, or low levels of technological diffusion) without investigating how the labor market has evolved over the past decades. Therefore, the study proposes three factors that may have had unintended consequences. These factors could have distorted the labor market by creating labor market duality between the public and private sector, as well as within the private sector between formal and informal economies. These topics are discussed in the next chapter.

2.6 Conclusion

The main ideas discussed in this chapter can be summarized as follows:

- Saudi Arabia can still be considered as a rentier state that is largely dependent on revenue from natural resources.
- The public sector, as virtually the only source of Saudi employment, is a relatively large sector suffering from underemployment and low productivity.
- The private sector can be generally stratified into two parallel economies: one is formal, developed, and driven by large professional firms; the other is informal, underdeveloped, and driven by micro stores. The former is vitally scarce while the latter is dominant, resulting in a shortage of meaningful jobs, lack of technological diffusion, poor working conditions, and over-dependence on foreign workers.
- Although they were low to start with, the average percentage contributions of the private sector to the GDP are likely to be overestimated because of its critical dependence on government purchases and subsidies and because of government joint ventures with the private sector.
- The structure of the Saudi industry and Saudi employment distribution reveals that the least localized (i.e., least Saudized) industries are the least concentrated while the most localized (i.e., most Saudized) industries are the most concentrated.
- The Saudi market structure is dominated by a high number of fragmented establishments (micro stores and small establishments constitute 97% of establishments), which is likely to hinder the transformation of the private sector into a developed and efficient sector.
- Saudis under the age of 30 constitute 60% of the population while the 15–29 age group accounts for 74% of the total unemployed local workforce.
- Unemployment costs are estimated at roughly 14% of the real GDP per year, equivalent to about USD95 billion (approximately SR356.25 billion), resulting from Saudi unemployment in 2016 only, let alone other social and political unemployment costs.
- There is a correlation between educational qualifications and unemployment rates, which contests the skill mismatch hypothesis as an explanation of unemployment persistence in Saudi Arabia.
- Proponents of the human capital theory as a solution for unemployment emphasize general training (educational reform) while seemingly neglecting factors such as

specific training, educated unemployment, and the lack of meaningful job opportunities for increasing labor force.

- Localization policies may be ineffective because of their limited impact only in the short term, and their counterproductive impact in the long term.
- Ineffective internship and apprenticeship programs contribute to the low attainment of experience and specific skills in the Saudi labor force.
- The Saudi labor force can be considered as semi-skilled. While middle-level jobs are scarce, Saudi workers are caught in the middle since they are unable to take over high-level jobs that require specific skills and are unwilling to take on low-level jobs that are abundant and occupied by foreign workers.
- The market size is relatively small and lacks investment incentives to create meaningful jobs. This is because of undefined property rights and high uncertainty pertaining to recurring changes in labor-related regulations and foreign labor supply.

3. Theoretical Framework of Saudi Local Unemployment: An Alternative Explanation

3.1 Introduction

According to Ethridge (2004), a theoretical framework should include the following:

> A) sources of the problem. This may address conditions, circumstances, policies, practices, etc., that cause(d) the problem. B) alternative solutions to the problem. C) Identification of variables relevant to the analysis of the problem. D) Conceptualized relationships in a system to analyze the problem. E) Hypotheses to be tested about results of analysis on the problem. (p. 130)

By adopting such a theoretical framework, parts A, B, and D are developed in this chapter while parts C and E are addressed in Chapter 4. This chapter discusses the evolution of the Saudi labor market over the last 50 years and its profound implications for labor market segmentation and Saudi (un)employment. The chapter is organized as follows: Section 1 examines three market-distorting factors: (i) the rapid expansion of the public sector and attraction of Saudi labor, (ii) the rapid expansion of micro stores and the disguised shortages of the national workforce, and (iii) the sponsorship system. Section 2 illustrates these factors by using supply and demand analysis, while Section 3 suggests an initiative to reverse or mitigate the consequences of these market-distorting factors.

3.2 Market-Distorting Factors and Structural Shift of the Saudi Labor Market

Structural unemployment uses the skill mismatch hypothesis as one of the explanations of unemployment when the gap between “skills acquired” (supplied) and “skills required” (demanded) widens in the labor market as a result of technological changes or structural shifts. This usually occurs because of technological advances, which prompt changes in the industrial structure, modifying the skills required. Thus, some skills become obsolete or in less demand, and part of the labor force becomes unemployed. However, Saudi Arabia has been experiencing an interesting phenomenon since the 1970s. In general, the Saudi labor force has become relatively skilled, yet the private sector has, by and large, failed to keep up with technological advances. Consequently, the public sector has become virtually the only source of Saudi recruitment while the private sector has become incapable of generating meaningful jobs for the increasing labor force. Hence, the present study argues that the Saudi labor market has become distorted by a structure that impedes its functionality. This is evident to such an extent that the implications of some labor theories (e.g., the human capital theory) and public programs to enhance the labor market have become inapplicable or ineffective. Thus, the Saudi labor market is characterized by policies and regulations that have had unintended consequences, creating disparity among different markets and labor immobility, ultimately leading to persistent unemployment. The structural shift of the market toward an inefficient economic state[11] can be attributed to three key market-distorting factors. The following three sub-sections elaborate on these factors.

[11] The inefficiency of the Saudi economy should be clear by now. In sum, the public sector is not only a saturated sector, suffering from underemployment but is also largely dependent on depleted and price-fluctuating natural revenues. Hence, it is susceptible to recurring economic shocks. In contrast, the private sector is impaired because of its critical dependence on governmental support and cheap foreign labor. At the same time, capital-intensive investment is limited, reflecting the low diffusion of technology. Moreover, the private sector has low participation and employment rates by the local labor force. Overall, the Saudi economy has been suffering from prolonged high unemployment, lack of economic diversification, and low resilience to prices fluctuations of natural resources.

3.2.1 Rapid Expansion of the Public Sector and the Attraction of Saudi Labor

The first factor that has contributed to the structural shift of the Saudi labor market is the rapid expansion of the public sector and the attraction of Saudi labor. Implementing the Five-Year Development Plans, particularly the first three plans (1970–1985), at an unprecedented rate since the 1970s has resulted in a rapid increase in public goods and services. This led to an urgent need for the government to attract Saudi workers. This was achieved by offering incentives, such as higher wages, regular working hours, training, and promotion, which were superior to those offered by the private sector. However, before proceeding further with the analysis, it is worth considering the Saudi government's employment policy.

The literature on employment decisions indicates that government employment decisions are made differently to those in the private sector (Demekas & Kontolemis, 2000; Ehrenberg & Schwarz, 1986; Freeman, 1986; Gelb et al., 1991). Hence, this field warrants closer examination. For instance, Demekas and Kontolemis (2000) maintain that government employment decisions have a significant impact on labor market performance because of two main factors.

The first factor is that "government actions—and particularly government employment policy—are dictated by the interests of the bureaucracy and the need to provide political favors to interest groups" (Demekas & Kontolemis, 2000, p. 392). Hence, using profit maximization models may be inappropriate to analyze government employment decisions (Ehrenberg & Schwarz, 1986). For instance, some authors attribute increasing Saudi employment in the public sector to the so-called "social contract" between the Royal family and citizens (Assaad, 2014; Assidmi & Wolgamuth, 2017). They highlight that "the basic terms of the contract are that rulers would provide citizens with oil revenues and citizens would provide allegiance, or political quiescence, to their rulers" (Herb,

2019, p. 5). Hertog (2016) points out that “the Saudi wealth distribution regime has contributed to the Kingdom's political stability for more than half a century” (p. 1); that is, since the Arab nationalist movement in the 1950s and 1960s. However, he maintains that such a distributive system “has been a key factor to shape (and distort) economy and labor markets” (p. 1). Further, conflicts of interest or favoring a particular group may seem obscure or irrelevant in the case of Saudi Arabia; however, some authors attribute the failure of the Saudization policy in the private sector to pressure from powerful sources. These include influential businessmen, whose personal wealth has been accumulated through cheap foreign labor, or the so-called “segmented clientelism” between different elites or different institutions (Hertog, 2006; Ramady, 2013; Spiess, 2010).

Since 1985, the government has implemented directive and indirective (normative) approaches, as embodied in its Five-Year Development Plans and the localization policies, to persuade the private sector to replace foreign workers with Saudis to alleviate unemployment. However, employers in the private sector claim that Saudi workers lack commitment or skills. Employers are also unwilling or unable to recruit and train Saudi workers at wage levels closer to those in the public sector. For instance, in a public meeting with businessmen during his term as Minister of Labor, Ghazi Al-Gosaibi summarized the issue of duality between the public and private sectors. His comments have been translated as follows:

> What does the young Saudi like about the public sector? The young Saudi likes job security, wages, reasonable hours, and reasonable days. I wonder if there is anyone among you [addressing the businessmen complaining about Saudi labor, with its high turnover rate and lack of commitment] who has spoken to one of his Saudi employees and said: O my son, I will let you work only five days a week, as in the public service, give you a wage as in the public service, and give you annual leave as in the public service. And then the employee leaves their job or flees from him? Impossible! (Al-Mutairi, 2015).

The second factor affecting government employment decisions is that “there is strong empirical evidence that the size of the government has a negative impact on overall growth performance …, as well as evidence that it has positive effects on unemployment persistence” (Demekas & Kontolemis, 2000, p. 392).

Therefore, because the government has played a vital role over the past 50 years in Saudi employment, this study considers the rapid expansion of the public sector and the consequent attraction of Saudi labor as one of the key factors to have distorted the structure of the labor market. Having discussed the importance of government employment decisions, we can now review the employment decisions of the public sector and explain how that might be related to the performance and duality of the Saudi labor market.

Over the period between 1970 and 2014, the Saudi economy has been oscillating between expansionary and contractionary fiscal policies reflecting oil price fluctuations. Generally, expansionary fiscal policy was adopted over the periods of 1970 to 1984 and 2005 to 2014 while contractionary fiscal policy was adopted over the period between 1985 to 2004. In the expansionary periods, Saudi employment increased, resulting from jobs created by new public projects and enlarging the size of government. In the contractionary period, Saudi employment also increased, though at a slower pace, because policymakers used Saudi recruitment in the public sector as a necessary measure in the face of economic and political challenges (Hertog, 2016).

Over the period between 1970 and 1984, increasing Saudi employment in the public sector was the main goal of decisionmakers while the private sector was neglected. Only from the fourth Five-Year Development Plan (1985–1990) onward, the government began urging the private sector to play a more active role in Saudi employment. By that time, the labor market was so highly segmented that the government policies aimed at increasing Saudi labor in the private sector—such as localization policies, government subsidies, and training programs—were largely ineffective. Referring to the first three Five-Year Development Plans, Niblock and Malik (2007) state the following:

> All three of the plans are characterised by their emphasis on the role of the state in shaping and leading the developmental process. None of them gives significant attention to the role of the private sector, and in fact none devotes a chapter to the sector. (p. 58)

Before 1984, the Saudi government did not encourage Saudi graduates to work in the private sector (Al-Asfour & Khan, 2014), seeking instead to attract Saudi labor to work in the public sector. Meanwhile, jobs in the private sector were progressively abandoned and were generally perceived as inferior to those in the public sector. Consequently, the compound annual growth rate of Saudi employment in the public sector between 1970 and 1984 was 7.71% (Saudi Arabian Monetary Authority, 2018).

Over the period between 1985 and 2004, Saudi recruitment in the public sector continued, but at a slower pace. Unlike the period between 1970 and 1984, there were no noticeable new projects undertaken to minimize the consequences of deteriorating oil prices at the time. A sort of institutional inertia prevailed at all levels. However, as the problem of unemployment became increasingly acute, with a growing labor force, limited jobs in the private sector as well as other social and political challenges (e.g., extremist groups), Saudi recruitment continued in the public sector. Consequently, the compound annual growth rate of Saudi employment in the public sector between 1985 and 2004 was 4.52% (Saudi Arabian Monetary Authority, 2018). By 1994/95, Saudi workers comprised 79% of the total employees of the public sector and only 10% of the private sector (Mahdi, 2000). At the time, the decisionmakers were in a difficult position, trying to balance the trade-offs between two decisions: increasing public sector employment as a strategy to reduce unemployment on the one hand, and reduce the budgetary deficit on the other.

Over the period between 2005 and 2014, because of the substantial financial reserves generated from high oil prices at the time, and because there had been a lack of projects across the country for the past 20 years (during the contractionary period of 1985–2004), the government felt compelled to resume its economic development. New projects were embarked in various sectors such as health, education, and even building so-called “economic cities” from scratch. Consequently, the compound annual growth rate of Saudi employment in the public sector between 2005–2014 was 5.65% (Saudi Arabian Monetary Authority, 2018). This period can be considered as the second economic boom.

Table 3.1 represents the compound annual growth rates of employment in the public sector for both Saudi and non-Saudi labor.

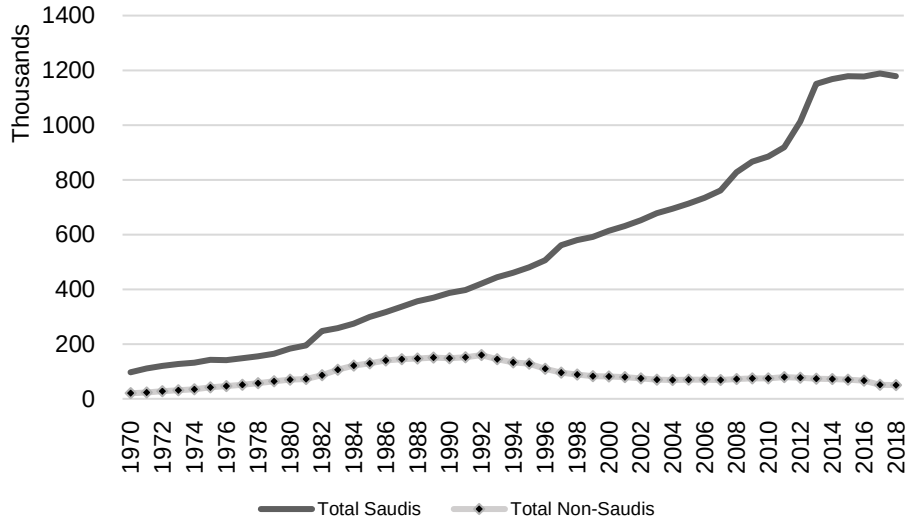
Table 3.1: Compound Annual Growth Rates of Employment in the Public Sector by Nationality

	1970–1979	1980–1989	1990–1999	2000–2009	2010–2018
Saudi	6.08%	8.07%	4.84%	3.92%	3.65%
Non-Saudi	13.66%	8.95%	- 6.23%	- 0.89%	- 5.03%

Source: The average annual growth rates were calculated by the researcher based on the number of employees in government sector (Saudi Arabian Monetary Authority, 2018)

As can be seen in Table 3.1, from the 1990s onwards, Saudi employment growth increased at a slower pace than in the 1970s and 1980s. This could have resulted from the replacement of non-Saudi labor rather than a response to new job creation. Figure 3.1 shows the number of Saudi and non-Saudi employees in the public sector over the period 1970–2018, with an annual average growth of 5.34% and 1.89%, respectively (Saudi Arabian Monetary Authority, 2018).

Figure 3.1: Number of Employees in the Government Sector by Nationality from 1970–2018



Source: Based on Saudi Arabian Monetary Authority (2018) data.

Figure 3.1 shows how the government had been using Saudi recruitment as a policy to respond to public sector expansion or political challenges. However, the government's eagerness to mitigate the issue of unemployment and income inequalities resulted in an increase of Saudi employment in the public sector and eventually, to underemployment and low productivity. When the public sector was noticeably saturated, public employment as a strategy to mitigate local unemployment reached its limit, and in turn, exacerbated the issue of Saudi unemployment.

The discussion now moves to the policy of rapid expansion in the public sector and the attraction of Saudi labor, and how this may create labor market duality between the public and private sectors. The following factors explain how the rapid expansion of the public sector and the attraction of Saudi labor have contributed to create labor immobility, and thereby, labor market duality and persistent unemployment.

First, the increase in the size of government and its role in the development process, particularly over the first three Five-Year Development Plans (1970–1984), has contributed to relegating the private sector to a shadow or informal sector. At the same time, attracting Saudi labor to the public sector while allowing the private sector to hire foreign workers has led many Saudis take for granted the idea of a job in the public sector. Hence, by the mid-1980s, labor market segmentation became prominent to such an extent, that the government implemented policies to increase Saudi labor in the private sector. However, these proved mostly ineffective. Consequently, the government had to continue using the Saudi employment policy for decades as a response to public sector expansion and/or as a necessary measure to mitigate unemployment. However, ultimately, this only served to exacerbate labor market duality between the public and private sectors.

Second, in the early stage of development, projects tend to be capital-intensive and long-term, which increases risk and uncertainty. At the same time,

while the private sector lacks incentives to take initiatives, public sector expansion has adopted a “project approach” rather than a “sectoral approach” (Niblock & Malik, 2007). This may explain why the private sector suffers from underinvestment, which has resulted in many economic activities being undeveloped or underdeveloped. Moreover, because the private sector is critically dependent on government purchases and subsidies (Ramady, 2013; J. Wright et al., 1996), and because of the absence of countercyclical policies (Wilson et al., 2012), the private sector remains vulnerable to oil price fluctuations. Hence, when the government adopts a procyclical policy or even reduces its subsidies to the private sector, businesses financially suffer or go bankrupt, exacerbating labor market issues such as job instability, low labor participation, and unemployment.

Third, while the public sector is ready to offer decent jobs to higher education graduates, there is a lack of professional firms in the private sector that can hire graduates with a vocational education. As a result, a negative stereotype is associated with vocational education, which is viewed as leading to unemployment or low-paid jobs. Moreover, increasing investments in higher education and expanding abroad scholarship programs might reinforce higher education over vocational training, thereby discouraging Saudis from taking up vocational education. Hence, the higher the level of education, the lower the likelihood of participating in the private sector because of the lack of suitable jobs (Kapiszewski, 2000; Ramady, 2010; Stevens, 1986).

Fourth, in the early phase of economic development, only a few cities were developed and urbanized, creating uneven regional development. Although there was political will to continue the development phase to other cities, the process stopped or slowed down because of the decline in oil revenues in the early 1980s. As a result, there was rural–urban migration, creating an imbalance distribution of population across cities and towns (Al-Filali & Gallarotti, 2012; Stevens, 1986; Wilson et al., 2012).

Fifth, there were crowding-out effects of government policies. One such effect stemmed from the policy of attracting Saudi labor to the private sector, especially during the first three Five-Year Development Plans from 1970 to 1984 (Al-Asfour & Khan, 2014; Birks & Sinclair, 1979; Stevens, 1986). This left the private sector lacking an adequate supply of national labor, and therefore having to depend on foreign workers. Another crowding-out effect stemmed from borrowing from domestic commercial banks during the period of 1985 to 2003 (Wilson et al., 2012). This reduced the ability of domestic financial institutions to lend money to private investors. Meanwhile, ceasing lending programs by the government-owned development funds also exacerbated the financial difficulties facing the private sector at the time (Al Hajjar & Presley, 1996).

Finally, while the private sector lagged in improving working conditions, the public sector adopted preferential policies for employment. For instance, there are two ministries for labor affairs:[12] the Ministry of Civil Service, concerned with labor in the public sector, and the Ministry of Labor and Social Development, concerned with labor in the private sector. There are also two systems of pension funds—the Public Pension Agency, concerned with workers in the public sector, and the General Organization for Social Insurance, concerned (generally) with workers in the private sector. Undoubtedly, this kind of structure sends a clear signal that there are wage differentials as well as other fringe benefits between the public and private sectors.

In short, because the characteristics of the public sector have improved while those of the private sector are still lagging in terms of adjustment and development, the public sector remains, even in times of austerity, the first-and-last-resort employer of the Saudi labor force. This becomes more apparent with the rise in Saudi employment as a policy in response to public expansion as well as other economic and political challenges. At the same time, the Saudi labor

[12] In 1/7/1441 AH corresponding to 25/2/2020, the Royal Decree number A/455 was issued, stating that the Ministry of Civil Service was to be joined under the Ministry of Labor and Social Development, with the latter name being modified to The Ministry of Human Resources and Social Development (Saudi Press Agency, 2020).

force still prefers the public sector over the private sector. In other words, the preferences of labor demand and supply in the public sector, to a greater extent, do match. In addition, when the public sector becomes saturated, Saudi recruitment becomes less effective or limited; hence, the unemployment issue among the local labor force becomes more acute. These reasons, among others, have contributed to sectoral differences and changed work incentives and preferences, resulting in occupational and geographical labor immobility. Therefore, the rapid expansion in the public sector and attraction of the Saudi labor have contributed to sectoral preferences, perpetuating labor market duality between the public and private sectors, and ultimately, contributing to unemployment persistence among the Saudi labor force.

3.2.2 Rapid Expansion of Micro Stores and Disguised Shortages in the National Workforce

The second factor that has contributed to the structural shift in the Saudi labor market is the rapid expansion of micro stores. Since the 1970s, the rapid growth of cities, in conjunction with the urban planning system which allowed building an excessive number of micro stores, created disguised shortages of the Saudi labor force, leading to the emergence of informal/secondary labor markets. Before elaborating on the issue of micro stores, it is important to clarify that this study is not opposed to small businesses per se. The objection here pertains to the status quo of the micro stores, whereby excessive micro stores were allowed to proliferate as opposed to, for example, designated shopping centers. In such centers, their numbers and positioning would have been carefully planned while restricting work for Saudi labor (this is discussed further in section 3.4.2).

The shortage of the national labor force has emerged for two key reasons. The first is the megaprojects undertaken in the 1970s, followed by the expansion of the public sector—all of which occurred over a short period. This type of labor shortage was mostly temporary because the labor needed to build these projects was easily overcome by foreign companies that won government contracts to

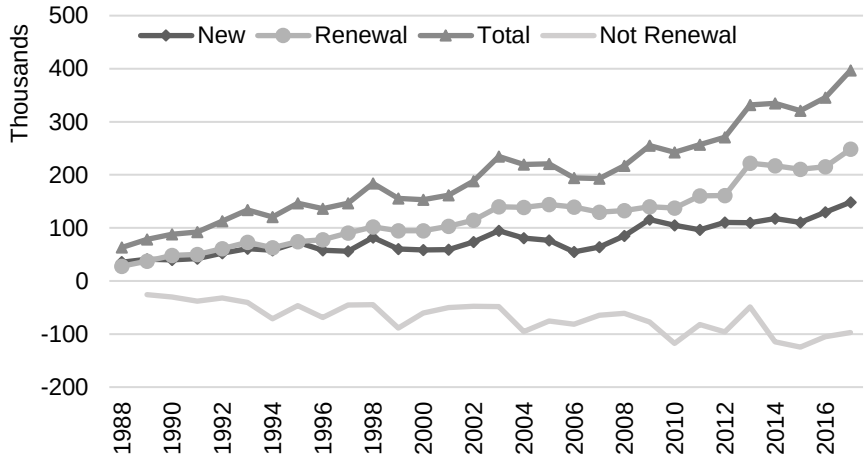
build the projects and were permitted to bring in unlimited foreign workers and raw materials. Although there was a need for foreign workers to operate labor-intensive institutions (e.g., health and educational institutions), the government was largely successful in replacing them by attracting more and more Saudi labor over time, as explained earlier. This is a well-known fact, clearly established in the literature (Ramady, 2010; Sirageldin et al., 1984; J. W. Wright et al., 1996).

The second, but less obvious, reason for the shortage of the national labor force is labor demand. As this study argues, the labor demand emerged *artificially* from the increasing micro stores. Because small enterprises are usually considered as a significant employer of a large percentage of the labor force, the micro stores model may be erroneously treated as small enterprises, especially by researchers unfamiliar with the Saudi labor market. To elaborate, the following factors shed light on how the rapid expansion of micro stores contributed to *disguised* shortages of the national labor force and to the emergence of the informal/secondary labor markets.

First, micro stores may generate a disguised shortage of the national labor force because increasing micro stores requires more foreign workers in a vicious cycle process. That is, as the number of micro stores increases, foreign labor also increases, which in turn partially increases micro stores even further, to meet the various needs of foreign workers. Existing urban planning exacerbates the issue by allowing micro stores to be built along the so-called “commercial streets.” Commercial streets differ from “high streets” or strip malls, which are restricted in designated commercial zones. In Saudi Arabia, commercial streets exist along the grid of the entire city, making many micro stores redundant. Because Saudi people do not usually work in micro stores and because the sponsorship system facilitates the influx of guest workers to meet the “labor shortage,” micro stores have become the prevailing business model, which is operated almost solely by foreign workers.

Currently, micro stores constitute 83% of the total establishments in the private sector (General Authority for Statistics, 2018). Moreover, micro stores grow at higher rate than the Saudi labor force. For instance, Figure 3.2 demonstrates the number of permits for establishments by different status from 1988 to 2017.

Figure 3.2: Number of Permits for Establishments by Status from 1988–2017



Source: Based on Ministry of Municipal and Rural Affairs (2018) data.

The compound annual growth rate of shop entry was around 6.54% (Ministry of Municipal and Rural Affairs, 2018). In contrast, the compound annual growth rate of the total Saudi population for the same period was around 2.26% (Saudi Arabian Monetary Authority, 2018). The fact that the growth rates of micro stores are higher than those of the Saudi population is likely to create artificial shortages of the national labor force. As discussed in Chapter 2, the continuous, increasing dependence on foreign workers is clearly linked to the rapid growth in the number of micro and small establishments, although most micro stores and foreign workers could be considered superfluous (Sadi & Henderson, 2005).

Second, the micro stores model can create a path-dependent situation. This is because policies aimed at improving industrial standards, implementing Saudization, or preventing “commercial concealment” are likely to be ineffective as long as the micro stores model exists. For instance, the government has

enacted several programs to assist the private sector to recruit and train national labor, including bearing a proportion of Saudi wages, which sometimes lasts up to a year. However, such programs failed, especially in small businesses.

Further, the dominance of micro stores in the private sector may be at the expense of a more productive, sustainable business model. According to the Annual Economic Establishments Survey (General Authority for Statistics, 2018), the relative distribution of small, medium, and big establishments is 83%, 14%, and 3%, respectively. These categories are defined by the number of employees: small/micro stores (less than 5), medium (5 to 19), and large (20+). If we assume that large firms create the good jobs prevailing in the primary labor market, then only 3% of the total establishments can be considered as creating meaningful jobs, which reveals how limited good jobs are in the private sector. In addition, the private sector is critically dependent on governmental purchases and subsidies, which in turn, is dependent on fluctuating oil revenues (Ramady, 2013). Hence, the private sector is not resilient to economic shocks. The prevalence of micro stores may have led the private sector to be largely peripheral and unreliable for delivering big projects. It is well known that the government seeks foreign firms for delivering most projects. In addition, big firms—both Saudi and foreign—are unable or find it difficult to do subcontracting work with Saudi small and medium establishments to deliver specific tasks because of the recurring issue of undelivered (failed) projects by those small and medium firms.

Third, micro stores create a disguised shortage of national labor because Saudi owners do not usually work in these stores. Instead, they tend to use them as a passive income tool by subleasing stores to foreign workers (i.e., the “commercial concealment” issue). If we pose the question of whether the micro stores model creates demand for Saudi labor, then the answer is definitively “no,” because of the “incentive incompatibility” between employers and employees. That is, while Saudi labor does not work nor have the willingness to work in micro stores, similarly, Saudi employers do not recruit nor have the willingness to hire Saudi labor either. Unlike the public sector, the preferences of labor demand and

supply in the private sector, to a greater extent, do not match. This makes the elasticity of (demand and supply) labor extremely low or in some cases, perfectly inelastic. Hence, voluntary transactions between labor parties are inconceivable in this case, while enforcing the localization policy would result in counterproductive effects.

Moreover, economists indicate that labor demand is a derived demand. In other words, different types of labor demand are induced by different markets, each with their own unique characteristics, and thereby, different labor requirements. Micro stores allow the emergence of a market that produces goods and services in poor working conditions at low wage levels. Consequently, the jobs that are created by this model have negative characteristics that do not usually match the desired characteristics and preferences of Saudi workers. In contrast, because big firms produce goods and services under good working conditions and create jobs with positive characteristics, many Saudi workers work or are willing to work for such firms. Thus, while the primary market creates satisfactory jobs, the secondary labor market creates unsatisfactory ones. Because most jobs in the private sector are created by micro stores and perceived as unsatisfactory, the private sector experiences low employment and participation rates of Saudi labor. After all, one may begin to wonder what type of business model has been unable to find enough labor for almost 50 years despite an increasing number of relatively skilled labor? There may be something wrong with the business model itself.

Fourth, the micro stores model, as a model almost solely dependent on foreign workers, contributes to depressing wages, reducing returns on human capital, delaying integrating national labor force into the private sector, fostering a consumption-oriented society, and perpetuating the problem of the low productivity of the private sector (Looney, 1988; Ramady, 2013; Sirageldin et al., 1984). According to Müller (2003), when the labor market is characterized by “good” and “bad” jobs, foreign workers will affect not only factors of production but also employment opportunities for the local labor force. As discussed in Chapter 2, the descriptive analysis of the private sector shows that the wage

levels of Saudi and non-Saudi labor differ, which leads to a preference among private sector employers for recruiting non-Saudi labor rather than Saudis. Hence, this continuous reliance on foreign labor may perpetuate low wage levels in the private sector. Further, increasing Saudi labor participation in the private sector and developing human capital are both difficult because of the absence of dedicated on-the-job training programs, including internship and apprenticeship programs. Hence, the micro stores model can be considered one reason for the de-skilling of Saudi workers because of the increasing dependency on foreign workers while Saudi graduates are unable to find jobs in the private sector. Another point is that the dependence on foreign workers has contributed to the vanishing of some artisanal trades or crafts. In the past, young Saudis would work as artisans, following in the footsteps of their fathers. However, with the increasing dependence on foreign workers, the Saudi youth tend to seek regular jobs. However, there may be more relevant socioeconomic factors that have played a role in the vanishing of artisanal trades, but this is a topic that falls outside the scope of this study.

Fifth, the micro stores model can be considered an unstable model. For instance, the average annual bankruptcy rate of shops between 1988 and 2017 was 36.23% (Ministry of Municipal and Rural Affairs, 2018), reflecting the instability of this model. At the same time, there are recurring changes in business owners or foreign workers, further accentuating the instability of the model. Moreover, Al-Hajjar and Presley (1992) indicate that 96% of 597 small business owners did not start up their businesses according to a business plan or feasibility study. This was attributed to three key reasons: a lack of data, a lack of government regulations, and the fact that such businesses are too small to require a feasibility study.

Moreover, jobs created by micro stores, such as shopkeepers or manual laborers, do not have advanced career prospects and are considered “dead-end” jobs. Such jobs do not offer promotions or offer good wages near the average Saudi reservation wage, making them non-promising and undesirable. Such jobs also have a non-motivating environment and poor working conditions. Unlike

climate-controlled or self-contained shopping centers, micro stores extend arbitrarily and exist everywhere, in urban and rural areas as well as near and far from the cities. Hence, they are susceptible to the harsh climate that regularly affects the country such as extreme heat and dust storms, discouraging Saudi labor participation, particularly among women, who feel insecure or uncomfortable. Through such negative working conditions, micro stores exacerbate sectoral employment preferences. Attracting national labor to the public sector while leaving many private sector activities underdeveloped contributes to shaping sectoral employment preferences of national labor, which favors administrative over labor-intensive jobs; hence, the low rates of participation and employment in the private sector. In other words, the undesirable characteristics of micro stores limit Saudi labor participation in this market. And if Saudis do participate, this is usually temporary, leading to high labor turnover and even higher costs of Saudi recruitment.

Finally, the micro stores model can be considered an inefficient model slowing down transformation into a more developed economy. In addition to the previous factors, the micro stores model is inefficient because of the limited capacity of the stores to utilize economies of scale and scope. In other words, the average cost of certain goods or services could be reduced if production volume were to increase or if they were produced along with other goods. Economies of scale and scope can be successfully utilized by large firms or through vertical and horizontal integrations, both of which cannot be achieved using the micro stores model. Having discussed the negative effects of micro stores, one may wonder how the government could mitigate the effect of micro stores and what alternatives could be offered. This important question is answered in sub-section 3.4.2.

In short, while the good working conditions in the public sector and the poor working conditions in the private sector have created labor market duality between these two sectors, the limited availability of good jobs in large professional firms and the plentiful but undesirable jobs in micro stores have created labor market duality in the private sector between the primary and

secondary labor markets. Moreover, while the labor shortage in the public sector was genuine, temporary, and overcome by attracting Saudi labor, the labor shortage in the private sector is continuous and mostly inflated artificially, by increasing the number of micro stores. Finally, while the public sector has expanded by essential infrastructure and Saudi employment, the private sector has expanded mostly by redundant micro stores, dependent almost solely on foreign workers. Hence, treating the micro store model prevailing in Saudi Arabia as a small enterprise model may lead to serious consequences in terms of policy implications and income inequality. The above characteristics of micro stores result in occupational and geographical factors, reducing or even preventing labor mobility by discouraging the Saudi labor force from participating or moving within the private sector between the formal and informal labor markets. Therefore, it is argued that the rapid expansion of micro stores has contributed to the emergence of the secondary labor market, perpetuating labor market duality between the formal (primary) and informal (secondary) labor markets within the private sector, and ultimately contributing to the persistent unemployment among the Saudi labor force.

3.2.3 Sponsorship (*Kafala*) System

The third factor that has contributed to the structural shift in the Saudi labor market is the sponsorship system. This system is embodied in Saudi labor law,[13] which regulates labor issues pertaining to foreign labor in the Saudi labor market. The sponsorship system facilitates the influx of guest workers into Saudi Arabia as employees. In the 1970s, the unprecedented rate of government spending on megaprojects resulted in a huge labor shortage, both in the public and the private sector. To mitigate this shortage and facilitate business opportunities for investors, the government relaxed the issuance of work visas for investors to bring in foreign workers. This was on the proviso that the investors were holders of business licenses and were able to sponsor (i.e., be the guarantors of) foreign

[13] See the Saudi Labor Law, Chapter 3, Articles 32–41.

workers, in compliance with Saudi labor law—hence the term “sponsor.” In other words, guest workers had to have a Saudi sponsor to be able to enter Saudi Arabia, obtain a temporary residence card (*Iqama*), and work for the sponsor. Although they can renew their residence card on a yearly basis as long as they have a valid sponsorship contract, foreign workers cannot obtain permanent residence nor can they be treated as permanent immigrants, irrespective of how long they have resided in Saudi Arabia. When foreign workers wish to work as freelancers, they must apply for a different type of visa before entering Saudi Arabia. This is usually difficult because of the financial situation of many foreign workers. If the sponsor no longer requires the guest worker, they either transfer or terminate the sponsorship contract through legal channels. This process is not always followed, as is illustrated shortly.

The sponsorship system was supposed to be a short-term, flexible policy in response to national labor constraints in the 1970s. However, it did not take long for sponsorship to become symbolic, as the sponsors’ role morphed into that of middleman, enabling foreign workers to enter the KSA. Moreover, the sponsorship system has several loopholes, distorting the labor market mechanism as well as jeopardizing its reputation. The following discussion examines how the sponsorship system reinforces labor market duality between the primary and secondary labor markets as well as other distortive effects on the Saudi labor market.

First, the sponsorship system may lead to the problem of adverse selection and moral hazard resulting from asymmetric information among labor parties. The issuer of work visas is unable to differentiate between “real investors,” who engage in genuine production but need foreign labor to fulfill the shortage of local labor, from “passive income seekers,” who wish to bring in foreign workers to let them work as freelancers while receiving monthly payments in return. Even with stringent restrictions, the system cannot control the behavior of sponsors or foreign workers. For instance, some sponsors use micro stores, which are accessible at low cost, to circumvent restrictions on foreign workers’ visas. Other firms abuse the government’s permission of obtaining unlimited work visas when

winning government contracts. Because they usually recruit more foreign workers than they need, some firms, after delivering the project, tend to illegally transfer the sponsorship contract or simply free up foreign workers, allowing them to work in the market for a monthly payment.

In contrast, when foreign workers come to Saudi Arabia and start working for the sponsor, some of them engage in opportunistic behavior to indirectly coerce the sponsor to let them work as freelancers. Moreover, although sponsors bear the costs for bringing in guest workers, they usually do have insufficient information about the prospective foreign workers, such as their skill level, commitment, and whether they would engage in opportunistic behavior after coming to Saudi Arabia. By the same token, foreign workers must also bear the costs of coming to Saudi Arabia, and may also lack information about their prospective sponsor and their compliance with the labor law. Because there is no effective method of mitigating such issues, and because of the casual labor relationship in the private sector owing to the prevalence of micro and small establishments, these problems—as well as other issues arising from the asymmetric information of the sponsorship system—tend to result in antagonistic labor relationships, increasing the risk and uncertainty of the business environment of the Saudi labor market.

Second, the sponsorship system may impede the free market mechanism by increasing labor frictions and inflexibility, which exacerbates labor immobility. For instance, restricting the right of transferring sponsorship contracts upon official approval or the sponsor’s consent may decrease foreign labor mobility. Empowering the sponsors over foreign workers may also lead to the exploitation of guest workers (Mahdi, 2000). In addition, the decision to determine the number of work visas could be considered subjective. The issuer of work visas may not always distinguish between large versus small businesses or between skilled versus unskilled workers. When the government wishes to enforce localization policies, work visas become more restricted; however, these restrictions may be arbitrary or ineffective. For instance, different investors with similar situations may be granted different numbers of work visas. With fewer restrictions on work

visas, some of the sponsors tend to apply for more than they need as a strategy for “labor hoarding” or more euphemistically, “labor stabilizing” (Ramady, 2010). Some sponsors apply for extra work visas during the less restricted period, then transfer (sell) the contracts during the restricted period. Moreover, the current system can also create conflicts of interest. This is because the government’s goal is to decrease the over-dependence on guest workers whereas the goal of private recruitment agencies (intermediaries) to whom the arrangement of bringing guest workers is delegated, is to maximize brokerage fees or profits by increasing the number of foreign workers.

Third, the sponsorship system may impede economic growth by discouraging foreign and domestic investors to invest in the Saudi economy. The recurrent changes in legislation pertaining to the issuance of work visas and the localization policies increase the risk and uncertainty, which may curb market growth and slow the economy, exacerbating the faults in the private sector (International Monetary Fund, 2018; Kapiszewski, 2000). Even with complete due diligence, investors may not be able to ensure stable legislation, making the investment in the Saudi market riskier than in some other regional or international markets.

Fourth, the sponsorship system, combined with the excessive number of micro stores, may contribute to the over-dependence on low-skilled foreign workers. The system permits any owner of a micro store to apply for work visas, entitling them to bring in guest workers. As discussed earlier, the excessive number of micro stores contributes to the influx of foreign workers, which does not necessarily lead to positive outcomes. Some have argued that increasing foreign workers does not have adverse effects (Harry, 2007). However, the problem does not lie in the existence of foreign workers per se, but in the increasing dependence on them at a time while high unemployment persists among local labor. Besides, the increasing influx of foreign workers—particularly that stemming from the micro stores, which create disguised shortages of national labor—can lead to several adverse effects. For instance, many foreign workers are low- or semi-skilled; thus, they are unlikely to play an important role

in knowledge spillover (i.e., transferring expertise). Several authors have asserted that the over-dependence on foreign workers contributes to various issues, such as low returns on human capital or slowing the development of human capital, depressing wages, and hindering the integration of the national workforce into the labor market (Al-Dosary, 1991, 2004; Looney, 1991; Ramady, 2013; Sirageldin et al., 1984; J. W. Wright et al., 1996). Sadi and Henderson (2005) state that many of the foreign workers residing in Saudi Arabia can be regarded as superfluous while Al-Dosary (2004) cautions against the various problems arising from the presence of foreign workers, including economic, social, psychological, and security issues. Hence, we can cite Gresham's law here with similar reasoning and state that "unskilled workers drive out skilled" and "bad business drives out good." This may be considered as an example of the adverse selection problem. Of course, this should not be taken out of the context and be construed as a xenophobic view or focus only on the negative effects of foreign workers. On the contrary, utilizing foreign labor has several benefits, including the ability of the economy to produce certain goods and services more effectively in terms of time and costs (J. W. Wright et al., 1996). Certain categories of foreign workers are also considered complementary to Saudi workers, which is important to consider when enforcing the localization policies aiming at increasing Saudi employment. The objection here is to the adoption of a model that contributes to excessive foreign workers while a large proportion of national labor goes without a job.

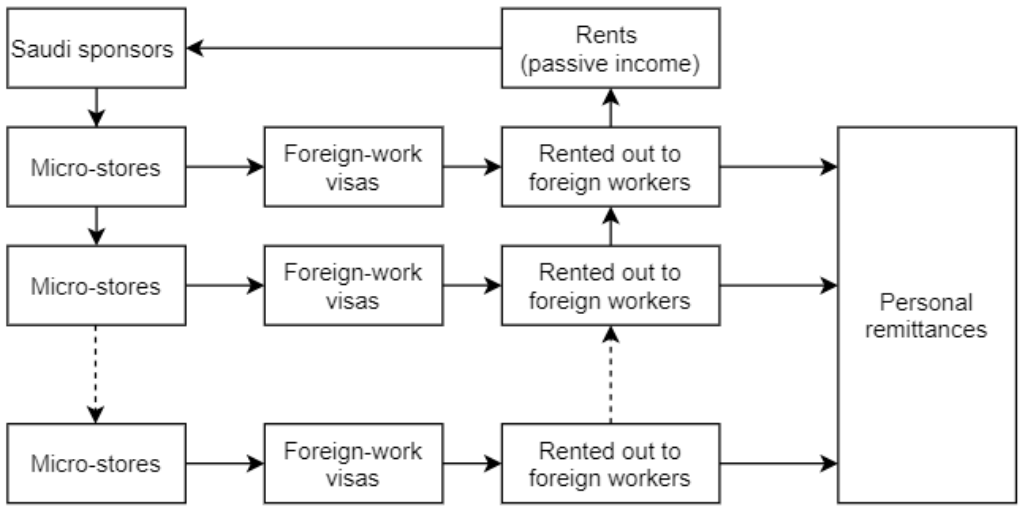
Finally, the sponsorship system, combined with the excessive number of micro stores, may contribute to so-called "commercial concealment." This phenomenon occurs when sponsors (Saudi or foreign investors) allow guest workers to work as freelancers by subleasing micro stores or letting them work freely in the labor market for monthly payments in return. Such a practice results from one or more of the following factors: (i) because Saudi employers, particularly owners of micro stores, do not usually run their business on a day-to-day basis, this makes monitoring foreign employees difficult; (ii) the temporary cessation or restrictions on issuing work visas or the inability to transfer the "sponsorship contract" reduces foreign labor mobility; (iii) because of the sunk

costs paid by Saudi employers to bring in the guest workers, some guest workers engage in opportunistic behavior to indirectly coerce Saudi employers to sublease micro stores to foreign workers; and (iv) because of the low start-up costs of micro stores and the preference of Saudi employers (sponsors) to make additional income while working elsewhere, subleasing stores to foreign workers has become a passive income tool (i.e., rent-seeking activities), especially when jobs and wages in the public sector are restrained.

These arguments may conform to the model of firm size distribution developed by Lucas (1978), who devised a model explaining the relationship between the “talent” distribution and size distribution of firms. One implication of the model is that the average firm size is the ratio of employees to managers, which is an increasing function of the marginal manager. In other words, there exists a cut-off level (e.g., reservation wage), according to which a person will decide to become either an employee or a manager. Hence, when the wage level (as the opportunity cost of being a manager) declines, it encourages some people to start their own businesses (Lucas, 1978; You, 1995). Thus, because of the low wages in the private sector on the one hand, and the restrained wages in the public sector on the other, many Saudis prefer to start their own businesses (usually micro stores), rather than working for someone else in the private sector for a lower wage. Eventually, subleasing micro stores to foreign labor has become a win–win situation or a mutually beneficial transaction between Saudi sponsors and foreign workers.

Because of the nature of the contractual relationship between sponsors and foreign workers facilitated by the sponsorship system, commercial concealment is widely practiced. Figure 3.3 illustrates the loophole of the sponsorship system leading to commercial concealment using micro stores.

Figure 3.3: Micro Stores, the Sponsorship System, and the Loop of Commercial Concealment



Source: Compiled by the researcher

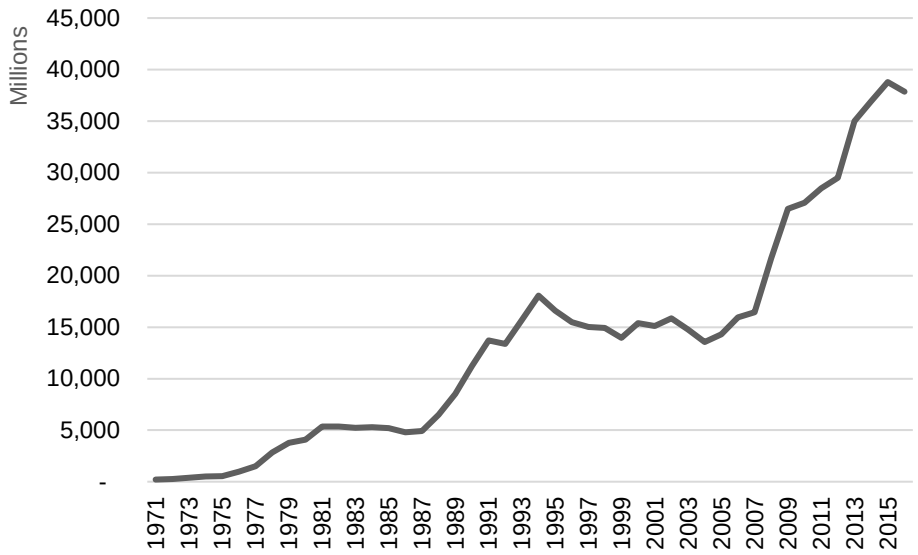
As demonstrated in Figure 3.3, first, Saudi[14] sponsors establish micro stores, then they apply for foreign visas. As business owners, they are usually entitled to receive at least one foreign work visa. Then, after the foreign workers arrive in Saudi Arabia, sooner or later the Saudi sponsors and foreign workers agree to engage in commercial concealment for a monthly rent paid to the sponsors while the rest of the money that foreign workers make is transferred outside the economy as personal remittances.

Several studies have cautioned against the negative effects of increasing levels of personal remittances. This is because, among other reasons, a large proportion of these may stem from the increasing activities of the underground economy, including commercial concealment (Al-Thaqafi, 2000; Kapiszewski, 2000; Sadi & Henderson, 2005; J. W. Wright et al., 1996). For instance, Figure 3.4 shows that the personal remittances transferred in current USD have

[14] Most of the sponsors are Saudis, but non-Saudi sponsors (i.e., foreign investors) may also engage in commercial concealment.

increased since 1971 at an average annual growth rate of 14%, ranking Saudi Arabia the second largest country in the world after the United States in transferred personal remittances (World Bank, 2018).

Figure 3.4: Personal Remittances, Paid (Current USD) from 1971 to 2016



Source: Based on World Bank (2018) data.

Because micro stores are the dominant business in the private sector, in which commercial concealment occurs (because of the virtually non-existent restrictions in such businesses), it is likely that a large portion of personal remittances is illegally generated through micro stores. In recent reports, commercial concealment was estimated to constitute up to 70% of micro and medium establishments ("100 billion volumes of commercial cover-up," 2017) while the Ministry of Commerce and Investment reveals that the size of commercial concealment ranges between SR300 to 400 billion annually in all industries ("The size of the commercial concealment," 2019). Although there could be some exaggeration or difficulty in estimating commercial concealment, like any other activity of the underground economy, the indications of this practice and its negative consequences have recently become too severe to be overlooked.

The government is striving to combat commercial concealment; however, their approach might be considered as treating the symptoms, and not the disease. The nature of the contractual relationship between sponsors and foreign workers, which is facilitated by the sponsorship system and the excessive number of micro stores, may unintentionally contribute to such unlawful practices. As mentioned earlier, micro stores create a path-dependent situation. This results in several issues; policy may not be effective in overcoming these issues unless the government reconsiders the regulation of micro stores in the first place. For instance, Al-Buraik (1988) points out the differences between individuals and institutional sponsors. The former refers to single sponsors who only become sponsors because they own a small shop (e.g., micro store) whereas the latter refers to any institution, either large or small, public or private, that becomes a sponsor of foreign workers. The author explains important differences between the two types of sponsor, including the fact that institutional sponsors tend to be relatively large and economically stable whereas individual sponsors tend to bring foreign workers to work as “partners” or simply free them in the labor market for monthly payments.

In short, despite the advantages of the sponsorship system that were achieved in the past, it could be argued that the advantages do not outweigh its adverse effects, particularly in the current complex and dynamic labor market. Indeed, it may contribute or exacerbate labor market distortion and labor frictions. First, the sponsorship system can lead to the problems of adverse selection and moral hazard because of asymmetric information among stakeholders. Second, the sponsorship system impedes the free market mechanism by increasing labor frictions and inflexibility, which exacerbates labor immobility. Third, the sponsorship system exacerbates the over-dependence on cheap and low-skilled foreign workers, creating a path dependence model. Finally, the sponsorship system, combined with the excessive number of micro stores, contributes to commercial concealment. Therefore, the government may need to consider abolishing the sponsorship system and replacing it by, for example, a centralized recruitment agency (discussed in section 3.4.1). Such an agency would allow the

government to internalize information, thereby mitigating several issues stemming from the problem of asymmetric information.

All in all, this thesis argues that those three policies (i.e., the market-distorting factors: allowing the rapid expansion of the public sector and the attraction of Saudi labor, allowing the rapid expansion of micro stores, and the sponsorship system) have unintended consequences; they distort the labor market, contribute to labor market duality and exacerbate labor immobility, thereby creating imbalances in labor demand and supply. The Saudi labor markets arguably face imbalances in demand and supply of labor, leading to labor shortages in some markets and excessive labor in others. The inability to find jobs in the public sector or the primary labor market, and the unwillingness to work in the secondary labor market, would both increase unemployment. Consequently, the Saudi economy has been experiencing this dilemma for several decades and has failed to create decent and promising jobs for citizens. Therefore, unemployment persists unabated.

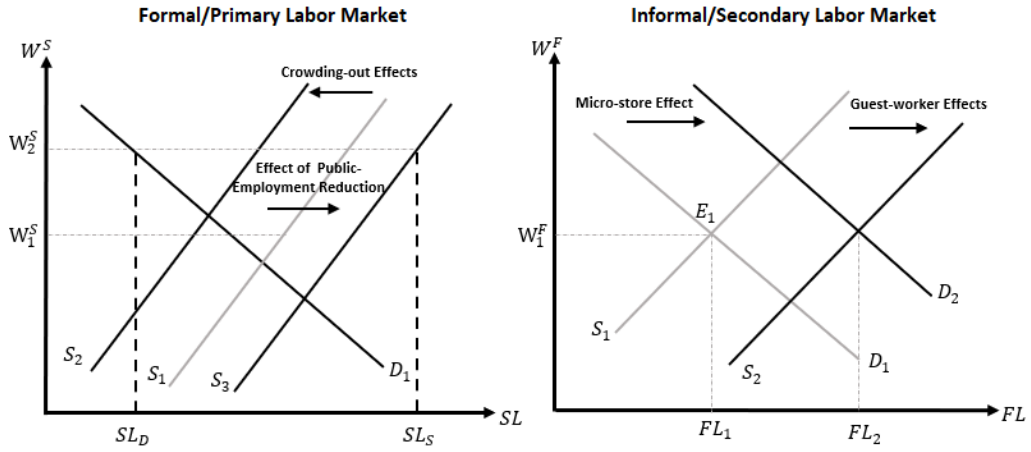
3.3 Effects of Market-Distorting Factors and Unemployment: A Graphical Analysis

This section examines all three market-distorting factors in combination to explain graphically how they have contributed to the labor market duality in the private sector over the last five decades and hence, their impact on labor issues. The name of each labor market is not of any great importance. The crux of the matter is understanding that there are two types of labor markets, each with working conditions and jobs that are diametrically opposed to the other. Consequently, they have profound implications for Saudi employment and participation as well as localization policies. The private sector can be generally stratified into two parallel economies: one is formal, developed, and driven by large professional firms (the primary labor market); the other is informal, underdeveloped, and driven by micro stores (the secondary labor market). The

former is extremely scarce, while the latter is exceedingly dominant, resulting in a dearth of meaningful jobs, lack of technological diffusion, negative working conditions, and over-dependence on foreign workers. In effect, the two labor markets have different conditions of labor demand and supply. Hence it is crucial to bear in mind those conditions when analyzing the Saudi labor market.

Figure 3.5 presents a hypothetical model of the dual labor market in the Saudi private sector and its impacts on the local unemployment.

Figure 3.5: Hypothetical Model of the Dual Labor Market of the Saudi Private Sector and Unemployment of the Local Population



Source: Compiled by the researcher

As can be seen in Figure 3.5, there are two labor markets: the primary labor market (left), and the secondary labor market (right). The former is characterized by the dominance of large professional firms and skilled and semi-skilled Saudi and non-Saudi labor (though the focus here is on Saudi labor). The latter is characterized by the dominance of micro stores and low-skilled foreign workers. Before the 1970s, let us assume that labor demand and supply were D_1 to S_1 in both labor markets and the levels of wage reservation in the primary and secondary labor markets were at W_1^S and W_1^F, respectively, with a low unemployment rate among the local workforce. But from the 1970s onward, the Saudi private sector dramatically changed. In the informal labor market, the rapid

expansion of micro stores resulted in a new dominant business model, shifting labor demand to the right (from D_1 to D_2), creating a disguised shortage of Saudi labor, termed in this thesis as the "micro store effect." The labor shortage prompted the government to relax work visa requirements through the sponsorship system, resulting in a shift of the foreign labor supply curve to the right[15] (from S_1 to S_2), known in this thesis as the "guest worker effect." Whereas the unemployment rate among foreign labor is effectively zero because they cannot enter Saudi Arabia unless they hold a work visa. The latest statistics show that the unemployment rate among foreign labor was around 0.8% (General Authority for Statistics, 2016).

In contrast, in the formal labor market, there were several factors that first caused the Saudi labor supply to decline (shifting from S_1 to S_2), and then increase (shifting from S_2 to S_3). First, the decline in Saudi labor supply was caused primarily by the rapid expansion of the public sector and the attraction of Saudi labor through higher wages and other fringe benefits. This is termed the "crowding-out effect" over the total supply of Saudi labor because of increasing government spending and employment policies between 1970 to 1984. Then, there was an increase in Saudi labor supply, caused primarily by boosting the Saudi workforce and the limited role of public employment policy. This was because the public sector became saturated and/or the efforts of the government to control the budget deficit between 1985 and 2004. However, the sluggish Saudi labor demand in the formal labor market can be attributed to the limited role of the private sector in creating meaningful jobs because of the market-distorting factors and their ramifications in the labor market, as explained in the previous section. Therefore, it can be argued that the limited role of the public sector in recent years as the main employer of the Saudi workforce, coupled with the limited role of the private sector in creating meaningful/primary jobs, have contributed to unemployment persistence among the Saudi workforce.

[15] Regardless of the wage level in the informal labor market, the shift must be to the right of E_1 to reflect the increasing number of guest workers.

Of course, the model is not meant to reflect actual labor elasticities nor the magnitude of the shifts in labor demand and supply. However, it is meant to dismantle the structural shifts that the Saudi labor market has undergone over the past 50 years, which have had profound implications for Saudi (un)employment and localization policies. Their impacts have shifted the Saudi private sector market into the dual labor market, restrained job creation, undermined the potential effectiveness of the localization policies, eventually leading to chronic high unemployment among the local population. Therefore, it is posited that these market-distorting factors play an imperative role in the analysis of the (un)employment levels in the Saudi labor market.

Before proceeding further, it is worth answering two questions that could be raised as an objection to the analysis. The first question is how the increasing wage levels in the public sector increased Saudi employment, and not unemployment, rates. The second question is how the public sector came to be the main engine of growth of Saudi employment in the 1970s and 1980s, later to become one source, among others, of unemployment. To answer these questions, we need to keep in mind the difference between static and dynamic economic analysis. Unlike static economic analysis, dynamic analysis allows a shift in labor demand and supply curves. Thus, in the 1970s, Saudi labor demand was greater than labor supply thanks to the rapid expansion of the public sector, which augmented the size of the government, followed by the attraction of the Saudi labor force, leading to Saudi employment growth in the public sector. This is not only consistent with the historical development of the Saudi economy but also with economic theory. For instance, economists have pointed out that government actions can increase wage levels while maintaining a high level of employment by shifting labor demand outward, instead of moving the quantity of employment along the demand curve derived from the production function (Freeman, 1986; Gelb et al., 1991).

To answer the second question, what happened was the opposite. While the Saudi labor supply increased as a result of the growing population and number of college graduates, the growth of economic sectors, including the public sector,

slackened as a result of economic recessions and the contractionary fiscal policy adopted at the time. Hence, Saudi labor supply was greater than labor demand. Consequently, because the average wage level in the public sector was, and still is, greater than that in the private sector, Saudi job seekers prefer to wait for vacant jobs in the public sector or in the primary labor market rather than working in the low-wage sector (i.e., the secondary labor market). This is consistent with economic theory. Assuming employees are risk neutral and there are two sectors—high-wage and low-wage—employees

> would move from the low-wage to the high-wage sector and remain … unemployed job seekers as long as the expected wage from choosing to wait exceed[ed] the expected wage of searching for work while employed in the low-wage sector. (Ehrenberg & Smith, 2012, p. 512).

In short, although the public sector is still the first-and-last-resort employer of Saudi labor, its effectiveness to absorb the increasing Saudi labor force has declined as the sector has become saturated in recent decades (International Monetary Fund, 1997). The public sector expansion and attraction of Saudi labor since the 1970s has resulted in an increase of Saudi employment, changing to underemployment in recent years. And because the demand for graduates of Islamic and managerial studies, as opposed to other specializations, is higher in the public sector, a considerable proportion of Saudi labor pursue such programs. However, when the public sector became saturated, educated unemployment increases among all specializations. In contrast, the private appears to be divided into two labor markets. The primary labor market has a limited role in generating jobs because of its small size, among other reasons, while the prevailing secondary labor market is incapable of creating jobs that Saudi job seekers would be willing to accept. Because of the dominance of micro stores and a lack of economic diversification, the private sector has largely become peripheral, lagging behind in development and diversity. Thus, the private sector is incapable of recruiting and training relatively skilled Saudi labor and relies instead on low-skilled cheap foreign labor. Thus, the inability to find jobs in the public sector or in the primary labor market, coupled with the

unwillingness to work in the secondary labor market, lead to high, persistent unemployment rates.

Therefore, it is posited that these three factors, which are the most important market-distorting factors, should be considered when analyzing the Saudi labor market. The last four decades have clearly shown the failure of the localization policies on the one hand, and the inability of the private sector to create enough jobs on the other. This negative outcome necessitates the government to take the lead by acting as a catalyst in the process of economic development. Against this backdrop, one may ask what the government can do to reverse or mitigate the distortion impacts of these factors. The answer to this question is provided in the next section.

3.4 Saudi Government's Role in Reversing the Effects of the Distorting Factors

In Saudi Arabia, some aspects of economic development could be described as "delusional prosperity" because of the unsustainability of the prevailing business model and increasing income inequality. In other words, it is likely that when oil reserves are depleted, a large portion of the private sector will vanish or deteriorate while most foreign workers will return to their home countries. Such an outcome would take the economy back to square one as a primitive economy. Then, the government and the private sector may not have the necessary means, in terms of financial and expertise requirements, to address the situation. To recall a less pessimistic scenario, when oil prices declined in 1983, the public sector entered a state of inertia, which lasted until 2005. During this period, the sector seemed unable to initiate new projects or even ensure the necessary maintenance of existing projects. The private sector, which is critically dependent on government purchases and subsidies, experienced numerous bankruptcies and the departure of a substantial number of foreign workers (Al Hajjar & Presley, 1996).

After reviewing many studies on issues affecting the Saudi labor market, it seems that there is no clear “exit mechanism” from the current dilemma facing the Saudi economy while at the same time maintaining market efficiency. Over the last 50 years, the size of government has increased through the undertaking of development projects, although this has had negative impacts on economic performance. The government has also attempted to implement both hard and soft policies to correct the labor market; however, these have proven to be ineffective. For instance, compulsory employment policies (e.g., localization, expatriate levy, etc.) have had counterproductive effects such as reducing the competitiveness of local businesses, inducing inflation, and ultimately, reducing job creation and employment retention. The private sector, for its part, has demonstrated impotence or apathy, unable to take the initiative and create decent jobs, despite constant support from the government. Meanwhile, the role of the public sector in Saudi employment became curtailed, resulting in underemployment and low productivity, not to mention the wage bill, which increased the budgetary burden. The aim of this section is to sketch a proposal to mitigate some of the issues affecting the Saudi labor market to improve the overall efficiency of the economy.

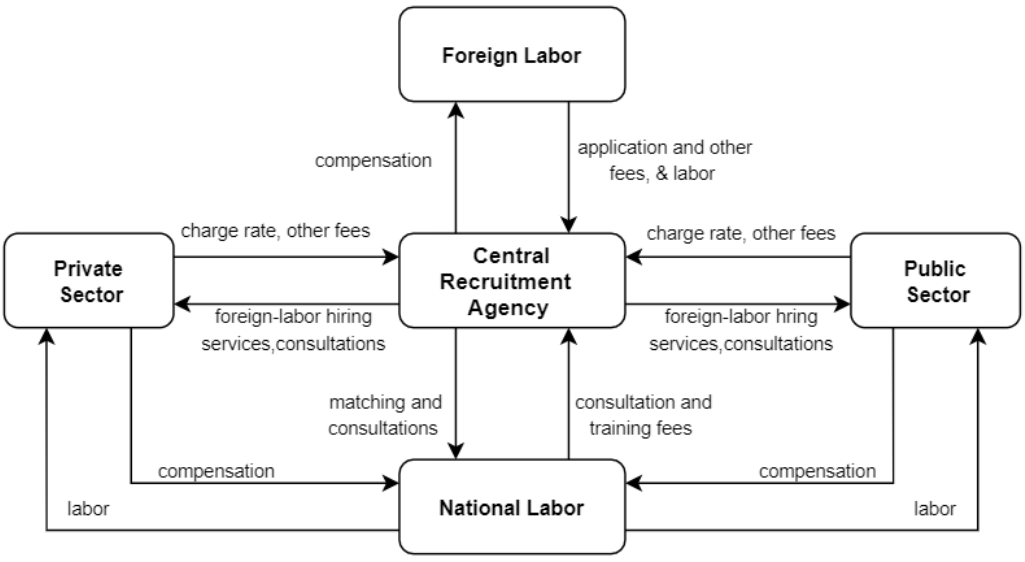
From the previous discussion, it is evident that the three policies have had unintentional consequences that have distorted the market mechanism of labor demand and supply, undermining economic efficiency. Overall, while the sponsorship system has distorted the total labor supply, the micro stores have distorted the product markets by adopting a low capital-intensive model and shifting the labor demand toward low-skilled cheap foreign workers. The public sector, with its appealing employment policies, has resulted in a sectoral preference to the detriment of the private sector. Hence, it can be argued that the economy requires a new market structure driven by market forces as a corrective measure for labor demand and supply. Because desperate times require desperate measures, the government may need to intervene initially by playing an active role as a catalyst and then leave the rest to market forces. Several corrective measures could be suggested to reverse the impact of market-distorting factors; however, two are singled out given the urgency of the

situation. First, the sponsorship system should be replaced by a central recruitment agency. Second, industrial restructuring should be undertaken to develop certain product markets or industries. The following discussion sheds light on these measures.

3.4.1 Central Recruitment Agency

If the government were to abolish the sponsorship system and replace it with a centralized recruitment agency (CRA) that would internalize information, it would be able to liberalize the labor market and enhance labor mobility. Possible labor relations between the CRA and other labor parties are presented in Figure 3.6.

Figure 3.6: Initial Visualization of Different Relationships Between the CRA and Other Labor Parties



Source: Compiled by the researcher

As shown in Figure 3.6, unifying the agency dealing with foreign labor has many economic, social, and political advantages. In terms of economic advantages, a key benefit would be the elimination of commercial concealment, which has

bedeviled the Saudi economy for many years. This is because the CRA would be the only employer of foreign workers, as an intermediate agency between foreign labor and local employers. Another economic advantage would be that the economy could reduce labor frictions and increase incentives for investors by reducing the uncertainty linked to foreign labor supply. For instance, the proposed CRA could use various methods to mitigate the problem of asymmetric information, including, for example, screening and assessment, labor credit system to incentivize employers and employees, skill classification, and improving job descriptions and contract design.

By internalizing information, the proposed CRA could reduce asymmetric information and thus enhance labor market efficiency, human capital allocation, and decision-making effectiveness. The tasks of the CRA could include the following:

- introducing national job classification and description that assist in designing training programs and improve skill acquisition;
- working as an intermediary institution between national labor and local employers, offering labor consultations, providing short labor-related courses, and improving job searches and matches;
- working as a monopsony employer of foreign workers and a supplier of foreign labor to local employers when there is a genuine shortage of national labor. In such cases, the CRA would charge local employers the average wage (certain charge rates). However, if the shortage was not genuine, the CRA would ask for the normal charge rate plus some sort of taxation. The tax, which could be considered as “externality tax” or a “targeted” expatriate levy, would be more effective than the current “general” expatriate levy because it would be more business-friendly and would have a less inflationary impact. To differentiate between genuine and non-genuine shortages, the CRA could use reliable, internalized information about the number of national workers registered at the CRA as job seekers, and possessing the skills required by local employers;
- integrating all or most other labor-related services, currently provided by different public agencies, into a single agency (i.e., the CRA). This would eliminate or reduce

resource waste and duplication as well as alleviating conflict of interest, thereby enhancing the efficient allocation of resources;

- creating a research center (think tank) on the development of the labor market, including the provision of labor statistics and periodic reports; and
- introducing a labor credit system (labor history) to incentivize employers and employees (national and foreign), which would foster compliance and commitment.

The benefits of the CRA can be summarized as follows:

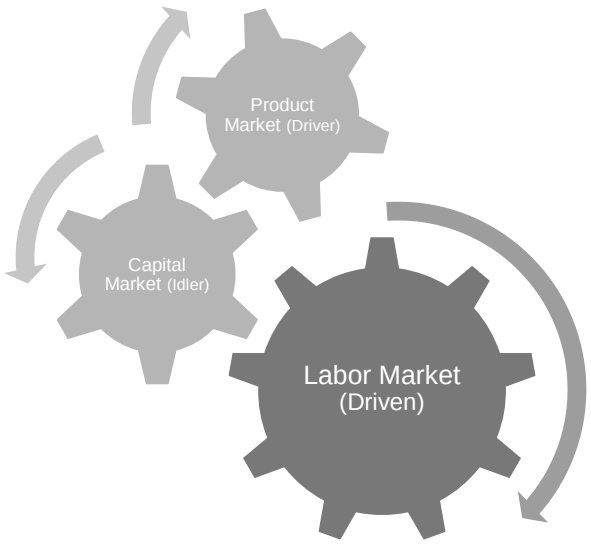
- eliminating the negative reputation of the sponsorship system;
- eradicating commercial concealment;
- reducing asymmetric information and labor frictions, which would increase labor mobility and curb unemployment;
- sharing relevant information and coordinating with the concerned institutions to improve education and training programs, national job classification and description, and professional licenses, which would create a level playing field and encourage labor participation;
- providing policymakers with more accurate and updated data to make informed decisions and improve labor institutions over time;
- playing a leading role in assisting both domestic and foreign investors by exercising due diligence or conducting appraisal studies before investing in Saudi Arabia which, in turn, would reduce risk and uncertainty. For instance, investors could consult the CRA, for a fee, to ensure beforehand whether the labor supply (national and foreign) is sufficient and its approximate cost;
- imposing higher charges for hiring *certain* foreign workers, which could be considered a "targeted" expatriate levy, and more effective than the "general" expatriate levy. A targeted expatriate levy would be more business-friendly and would have a less inflationary impact;
- generating revenues by providing consultations and foreign labor hiring services;
- reducing the negative effects of low-skilled foreign labor on the labor market, such as depressing wage levels;
- allowing frictional unemployment of foreign workers for a short term as it is considered more cost-effective than recruiting from abroad;
- providing complete information about the legal status of foreign workers, which would reduce their security issues, especially for those who are undocumented or are engaged in the underground economy;

- eliminating the restriction on foreign labor mobility by Saudi sponsors and thus bringing about market-driven labor mobility for both Saudis and non-Saudis, which would reduce the private employers' current preference for foreign workers because they are less mobile than Saudi workers;
- managing seasonal foreign labor (e.g., harvesting, pilgrimages) more effectively; and
- reducing the risk exposure of certain industries to foreign labor shortages, particularly for critical industries.

3.4.2 Industrial Restructuring

To reinforce the argument presented above, Figure 3.7 shows three gears, known as the driven, idler, and driver gears.

Figure 3.7: The Economy and its Three Gears



Source: The researcher

Figure 3.7 represents the three interconnected markets: the labor, the capital, and the product[16] markets. If the driver gear works, the other gears can work smoothly. But if the driver gear breaks down or is without teeth, the other gears

[16] Here, reference is made to the whole industry, not just to the final goods.

cannot move. The idler gear can be used to shift the direction of the other gears. This is similar to the economy, where labor demand is driven by the product demand, while the capital market, as a financial intermediary, channels funds to investments. It could be argued that the Saudi capital market is a functioning market while the product market, because of the dominance of micro stores, is a dysfunctional market (i.e., the driver gear is broken or without teeth). The market structure is dominated by micro stores; however, these cannot be considered as "serious" investments. Hence, despite the relative improvement in the supply of national labor, the demand for national labor in the private sector is still sluggish. Like the idler gear's function, the capital market can be used to shift funds to increase real investments and develop product markets, which brings us to the question of why there are not enough investments to increase the scale and scope of the product markets, which would induce the demand for national labor.

The basic premise here is that many product markets or industries—as still underdeveloped—require high fixed costs, which increases investment risk and time expected to bear profits. Besides, there are three main issues that exacerbate the problem of high fixed costs associated with the development of product markets. The first pertains to the relatively small size of the local market. Hence, the high fixed costs relative to the small market size resemble the situation of natural monopoly, while there is no entry restriction to reward first-mover firms to utilize economies of scale. The second issue pertains to the uncertainty surrounding foreign labor supply while local workers lack the specific skills or willingness to participate in the private sector. The third issue is related to the high costs of recruiting and training labor for industry- or firm-specific skills, which is typically a characteristic of the formal sector (i.e., the primary labor market). Labor economists consider labor as a quasi-fixed factor of production. Oi (1983) states the following:

> The discipline of labor economics has now accepted the proposition that labor is a quasi-fixed factor of production. … The fixed cost hypothesis was developed to explain the occupational differences in employment and wage rate responses to cyclical fluctuations in the aggregate level of output and employment. (p. 63)

Recruiting and training national labor can exacerbate the fixed costs, and works as a disincentive or even a barrier to entering the market. Therefore, these constraints are likely to be the main causes preventing the private sector from playing an important role in the development process. In other words, these constraints make private investors unable or unwilling to invest. Consequently, if the government cannot improve industries alone and the private sector is unable or unwilling to take initiative, what is the proposed model that could act as an "exit mechanism" from the current dilemma while ensuring market efficiency at the same time?

The proposed model is simply to undertake market/industrial restructuring, consisting of two parts. The first part would involve land re-zoning while the second would involve the development of product markets. Accordingly, the first part would restrict land use for commercial purposes only, and prevent the establishment of new micro stores by abolishing the existing urban system that allows micro stores to be built anywhere. Existing micro store owners could be encouraged to convert their stores into dwellings,[17] move into shopping centers, or become part of the new transformative large companies (explained below) as, for example, suppliers for department stores, warehouses or franchises.

Reducing the number of micro stores would decrease the redundant foreign labor while enhancing the profitability of small and medium enterprises, which in turn, would encourage local labor to participate in private sector activities. Designating specific geographical zones for shopping centers would not only improve profitability by utilizing internal economies of scale but would also improve working conditions, such as security and environmental issues, through external economies of scale. For instance, it would be easier to provide other supplementary services such as daycare centers or other amenities, which currently do not exist in the micro stores model. This would encourage the

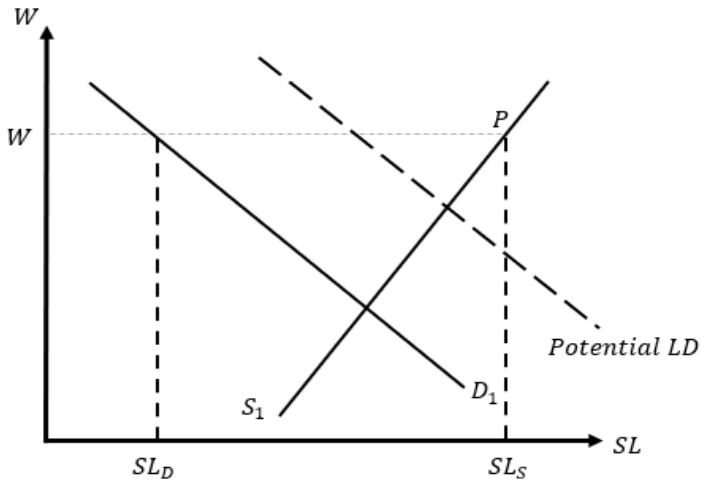
[17] Converting micro stores into dwellings would not be difficult because they are usually part of a dwelling building already. In fact, converting micro stores into dwellings is currently one of the requirements if a property owner wishes to add an additional level to the building. See some pictures of micro stores in Appendix 1 for better visualization of micro stores in Saudi Arabia.

national workforce, particularly women, to participate in the private sector. Internal and external economies of scale resulting from industrial clustering and geographical zoning are beyond the scope of this study, but are discussed comprehensively in the literature (Arnott & McMillen, 2006; Lipczynski et al., 2005).

The second part of the model would be to establish joint ventures with the private sector whereby these ventures would be granted government franchises, at least during a transition period, to develop certain industries or product markets. This would occur on the proviso that these companies adhere to certain guidelines, such as implementing professional standards and undertaking a decisive program for hiring and training national labor. The main goal would be to develop certain industries, where firms would be viewed as “transformative development companies (TDCs),” accelerating the transition into a more developed and efficient market. The basic role of government would be to design a “regulatory mechanism” for granting concessions for joint venture firms in each industry, and in return, the companies would be required to improve professional standards and recruit and train local labor. As a “clean slate” approach, the government would have the opportunity to set the guidelines and professional standards, thereby, improving working conditions and the quality of goods and services. As for employment policy, it would contribute to training and recruiting Saudi labor in the first place, while utilizing foreign workers whenever there was a real, not disguised, labor shortage. Although large firms would be unlikely to create disguised labor shortages, differentiating genuine from disguised labor shortages would be possible through the (proposed) central recruitment agency, as explained in the previous section.

Figure 3.8 illustrates hypothetically the total quantity of labor demand and supply in the Saudi economy, showing that the quantity of Saudi labor demanded (SL_D) is less than its total quantity supplied (SL_S), resulting in persistent unemployment.

Figure 3.8: Total Labor Demand and Supply in the Saudi Economy



Source: Compiled by the researcher

As can be seen in Figure 3.8, the going wage level could represent the minimum wage that Saudi labor would accept (i.e., the reservation wage), resulting in the quantity demanded of Saudi labor (SL_D) falling short of the quantity supplied (SL_S), which results in persistent unemployment. The existing demand for Saudi labor (the solid curve, D_1) is assumed to stem from the public sector and from the primary labor market in the private sector. The total demand for Saudi and non-Saudi labor is represented by the dashed curve (the potential LD curve). The difference between the total demand for labor and the existing demand for Saudi labor is currently offset by foreign labor. The difference cannot be attributed to the shortage of Saudi labor; otherwise, persistent unemployment would not exist. In fact, the difference is attributable to the dominance of the secondary labor market, making meaningful jobs scarce in the Saudi economy. In other words, a large portion of the product demand is being supplied by the secondary labor market, which recruits foreign labor. Hence, it is posited that the weak demand for Saudi labor cannot be attributed entirely to “deficient demand” because both aggregate demand and unemployment exist simultaneously, sometimes at high rates. Moreover, the weak demand for Saudi labor cannot be attributed entirely to a “skill mismatch” because most of the existing establishments are micro stores relying on unskilled or semi-skilled cheap foreign labor. Finally, the weak demand for Saudi labor cannot be attributed completely to frictional factors because Saudi unemployment is prolonged while frictional unemployment tends

to occur in the short term. Therefore, the persistent Saudi unemployment must be attributed to the limited number of professional large firms in the private sector that are capable of transforming the sector into a more formal, diversified economy while creating various meaningful jobs.

Therefore, the state could move the labor demand curve outward (e.g., to the potential LD curve) if the size of the formal labor market were to increase through real investments to improve the product markets currently supplied by the secondary labor market. In other words, the difference between existing labor demand and potential labor demand could be offset by transforming the informal market into a formal one. Hence, the proposed approach could accelerate the transition to the formal labor market. For instance, products and services, which are currently sold separately by different, fragmented micro stores, could be sold in specific department stores. In addition, the state could design a “regulatory mechanism,” through which several upstream and downstream businesses could be created (perhaps via joint ventures between the government and the private sector). This would expand the market size both horizontally and vertically, providing distribution channels and franchising opportunities for local small investors. This would encourage innovation and increase investment channels for local investors instead of relying on passive income micro stores. At the very least, it would create the necessary means for businesses to operate efficiently and be able to (as required by the regulatory mechanism) establish effective on-the-job training programs—a long overdue goal that the private sector has failed to deliver. Moreover, the TDCs could be assigned to develop self-contained shopping centers that provide supplementary services such as food courts or daycare centers. Because of the prohibitively high costs of land and lack of long-term leasing contracts (i.e., because of weakly defined property rights among other reasons), developing such shopping centers would require high fixed costs, which act as entry barriers for the private sector. But by assigning this task to the (proposed) TDCs while enhancing property rights, the size of the formal sector would expand, thereby increasing the demand for Saudi labor as well as enhancing efficacy, productivity, and national labor participation, particularly by women. This would productively employ Saudi men and women

in non-oil industries, leading to higher growth rates as well as other economic benefits.

Table 3.2 summarizes the main differences between the micro stores model and the large professional companies (the TDC model).

Table 3.2: Micro Stores Versus Large Professional Firms

Criteria	Micro Stores (Existing Model)	Large Professional Companies (Proposed TDC Model)
Business Intensity	Labor-intensive, usually cheap low-skilled labor	Capital-intensive, usually semi- and high-skilled labor
Technological Progress	Extremely low technological diffusion, incapable of driving development	High (or capable of) technological diffusion, capable of driving development
Economic Diversification	Extremely low level of scale and scope economies, non-sustainable business model	Large (or capable of) scale and scope of economies, sustainable business model
Working Conditions (e.g., job security, labor turnover, wage levels, advanced job opportunities)	Poor (e.g., wage below subsistence level, informal business)	Good (e.g., capable of recruiting Saudis at average wage level, formal/professional business)
Industrial Codes/Standards Compliance	Low (or unlikely)	High (or likely)
Labor Nationality and Participation	Almost exclusively dependent on cheap, low-skilled foreign labor with concomitant lack or low participation of local labor (usually confined to employer/manager/sponsor)	Usually higher- and lower-end jobs occupied by foreign labor, while middle jobs occupied by Saudi labor; high participation of local labor
Implications of Localization (Saudization) Policies	Low (or no) localization because of incentive incompatibility	High localization (or likely) because of incentive compatibility, especially for higher- and middle-job positions, although localization policies not recommended because of their adverse effects
Training Programs and Expertise Transfer	Low (or unlikely) internship and apprenticeship programs, inconceivable because of costs and other reasons	High (or likely) internship and apprenticeship programs conceivable by decreasing average costs

Source: Compiled by the researcher

As can be seen in Table 3.2, Saudi Arabia is at a crossroads to either continue adopting micro stores (the existing model) or large professional firms (proposed model) to transform its economy. If the government had adopted, in its early development plans, a different approach by allowing a few professional firms, greater chances could have been expected of creating better job opportunities, improving industry standards, and having a diversified economy. As discussed in section 2.2, the most localized and developed industries are the most concentrated ones. In contrast, micro stores are unable to utilize economies of scale and scope; hence, they are too small to adopt industrial standardization or recruit and train national labor, let alone drive the economy into a more diversified one. Because of the constraints preventing the private sector from playing an important role in the development process, granting concessions to a few companies, at least over a transition period, to develop certain industries and product markets would be a more efficient and effective approach.

Such an approach would be more efficient because of its ability to utilize economies of scale and scope as opposed to micro stores. It would also restore confidence in the local market and accelerate economic transition, as opposed to the existing soft and hard policies to develop the economy, most of which have failed. Such an approach would be non-disruptive to the market, unlike compulsory policies that usually lead to economic shocks and market disruptions. This is because the supervision body could prevent market disruption by allowing incumbent micro stores to continue their business as usual while the TDCs gradually penetrated different product markets. To ensure a quick yet smooth transition, micro store owners would have to choose among the above options when the TDCs cover certain occupational or geographical areas or after a specific date. Therefore, this approach would shift the private sector from an informal and underdeveloped one, driven by fragmented business entities (i.e., micro stores), into a formal and developed sector, driven by professional firms. TDCs could make a giant leap forward across the board while mitigating several persistent issues in the Saudi labor market. The advantages of the proposed TDC model can be summarized as follows:

- accelerating the transition to a more developed, sustainable, and diversified economy; hence, reducing dependency on natural resources and preparing the economy for the post-oil era;
- reducing dependence on low-skilled foreign workers while utilizing skilled foreign labor, which increases knowledge spillover and expertise transfer;
- because the private sector lacks expertise or interest in creating a business model that integrates local labor and increases efficiency, the proposed approach would raise the standards by setting an example for future entrant firms;
- enhancing wage levels and working conditions and normalizing work in vocational jobs, which would, in turn, increase participation and employment in the private sector as well as reducing unemployment;
- boosting the national competitiveness of domestic firms to compete both at regional and international levels;
- fostering the creation of industrial standards and standardization of some products, which would improve compliance, the quality of products and services, and training programs;
- by utilizing internal and external economies of scales, efficiency within (intra) and between (inter) firms would improve, which would enhance productivity, improve supply chain channels, create franchise opportunities, and mitigate the issue of subcontracting failure;
- improving internship and apprenticeship programs through training programs conducted by TDCs;
- instead of the fragmented establishments, integrating relevant goods and services into different business entities, which would increase the scope of the firms, improve efficiency, reduce transaction costs, and offer better quality of goods and services;
- the proposed TDC model is not only non-disruptive to the market but also more efficient and more effective in meeting development goals; and
- it would be easier to reach small and remote cities through the TDCs , which, in turn, would improve livelihoods in rural areas and small towns, thereby increasing labor mobility or at least reducing internal migration to main cities.

One may criticize the proposed TDC model and claim that it will cause some sort of supply restriction that would result in monopoly rent (abnormal profit), reducing employment and output. There are a number of responses to this criticism. First, the argument assumes that the Saudi labor market is competitive,

which is questionable. In fact, the Saudi labor market can be considered monopsonistic or imperfectly competitive for various reasons, one of which is the restriction on foreign labor mobility enforced by the sponsorship system (Feess, 2012). Second, many investments are neglected by private investors for various reasons, which can be considered as a market failure; hence, there are sufficient justifications for the government to intervene. Third, unlike enforcing localization policy, adopting the proposed model would have less, if any, market disruptions. This is because enforcing localization would increase costs to unwarranted levels or force firms to go bankrupt, as found by several studies (International Monetary Fund, 2018; Kapiszewski, 2000; Peck, 2017). Fourth, the internal and external economies of scale which would be derived from commercial zoning and concentrated industries would be likely to outweigh potential costs. For instance, large firms are likely to increase R&D spending and promote technological progress (Feess, 2012; Lipczynski et al., 2005) as well as offering higher wages and fringe benefits (Krueger & Summers, 1986; Manning, 2011). In addition, Demsetz (1973) cautions against reduced efficiency when implementing deconcentration or antimerger policies:

> The cost advantage that gives rise to increased concentration may be reflected in scale economies or in downward shifts in positively sloped marginal cost curves, or it may be reflected in better products which satisfy demand at a lower cost … Profit does not arise because the firm creates “artificial scarcity” through a reduction in its output. Nor does it arise because of collusion. Superior performance can be attributed to the combination of great uncertainty plus luck or atypical insight by the management of a firm … There is no reason to suppose that competitive behavior never yields monopoly power, although in many cases such power may be exercised not by creating entry barriers, but through the natural frictions and ignorance that characterize any real economy. (pp. 1–3)

Fifth, as mentioned earlier, the high fixed costs for developing certain industries relative to the local market size increase the long-term average costs, resembling the conditions of natural monopoly. Natural monopoly may be warranted when high average costs of a certain product or industry (e.g., public utilities) fall with an increasing level of output. However, the question may be posed, what causes the high average costs in the first place? Waterson (1987) maintains that:

> The traditional theory of natural monopoly implies that particular industries are concentrated, other not, because is in some cases fixed costs are high relative to the size of the market. Shaked and Sutton's insight is that in some cases, fixed costs are high because the market is large. This can be seen as a "demand pull" view of technical change: the size of the potential market is what calls forth product improvements. (p. 78)

Hence, average costs may increase because of high fixed costs relative to the small size of the market or resulting from the large size of the market that required high investment to improve product quality through technological progress. Either way, considering the small size of the Saudi local market or considering the need for developing product markets, several Saudi industries may warrant some degree of concentration or even monopoly. Even if we agreed with the objection that the proposed model might result in monopoly profits, some economists, "most notably J. A. Schumpeter, have stressed the beneficial role that monopoly profits can play in the process of economic development" (Nicholson & Snyder, 2008, p. 513).

Sixth, the fair practice of the proposed TDC model could be ensured not only by the supervisory body but also by partial government ownership (if present), which plays an important role in employment and production decisions, as discussed in Chapter 2. Finally, historical evidence reveals that the government has intervened by granting concessions to one or a few companies because of poor performance or to accelerate the development of certain industries and products. These have indeed resulted in, among other positive externalities, an improvement in working conditions and an increase in Saudi employment. For example, there were about 15 electricity companies before the Council of Ministers Order (No. 196) dated 11/08/1419H (30/11/1998), which stipulated the merger of all Saudi electricity companies into one consolidated company—the Saudi Electricity Company (Saudi Electricity Company, 2020). Another example is that there were many companies for gas refills and distribution, but because of poor or discontinued service, the government intervened and decided to merge all those companies with the National Gas and Industrialization Company at the end of 1975 (National Gas & Industrialization Company, 2020). Further, as explained in Chapter 2, SAMA has fostered several mergers between banks

and financial intermediaries to enhance the efficiency of the financial industry. Finally, the concession to the oil industry may be a good example to show how one company could improve the entire industry's efficiency while maintaining fair practice. However, this study has not considered the oil industry as the only example of a successful concentrated industry to avoid the argument that the oil industry is a special case as a sovereign sector.

Therefore, several industries arguably require restructuring by replacing the micro stores model through commercial re-zoning on the one hand, and by granting government franchises to certain TDCs to accelerate development and foster economic efficiency on the other. At the same time, the regulatory body would oversee market conditions and firm practices to ensure that firms conform to the guidelines and goals of developmental transition. After a certain transition period, the government may reassess market conditions and decide whether there is a need for more companies to enter the market.

Nonetheless, the suggested approaches may face some difficulties. As for the CRA, there are many private recruitment agencies currently operating in the economy that would lose their businesses if the government adopted the CRA model. Hence, adopting such a model would necessitate a proper mechanism to compensate their losses. Additionally, other difficulties may face TDCs, including how to distinguish between different industries in which the transformative firms would operate, what industries the government should start with, and how to devise a regulatory mechanism to hold transformative firms accountable for delivering on developmental goals.

3.5 Conclusion

The main ideas discussed in this chapter can be summarized as follows:

- In this chapter, three main policies were identified: (i) the rapid expansion of the public sector and attraction of Saudi labor, (ii) the rapid expansion of micro stores and disguised shortages in the national workforce, and (iii) the sponsorship system. These policies have unintended consequences on the Saudi labor market by distorting it, resulting in labor immobility and perpetuating unemployment.
- The impacts of the rapid expansion and attraction of Saudi labor in the public sector can be summarized as follows. First, adopting an employment policy in the public sector, either to attract Saudi labor or to curb unemployment, has contributed to sectoral preferences between the public and private sectors and to underemployment issues in the public sector. Second, slowing down the process of development and preventing it from reaching cities other than the main centers, because of oil price decline, has created an imbalanced distribution of population and geographical immobility of labor. Third, two crowding-out effects have stemmed from two policies, creating an adverse effect on Saudi employment in the private sector: (i) attracting Saudi labor from the private sector while allowing the private sector to hire guest workers may have created a labor market duality and a path dependence situation where the private sector becomes dependent on cheap foreign labor, impeding the integration of national labor into the private sector; and (ii) borrowing from domestic commercial banks in the late 1980s and 1990s reduced the ability of domestic financial institutions to lend money to private investors, thereby resulting in less economic growth and job creation. Therefore, it is argued that the above factors have contributed to sectoral differences and changed work incentives and preferences, resulting in occupational and geographical immobility of labor, and ultimately, unemployment persistence.
- The impacts of the rapid expansion of micro stores can be summarized as follows. First, while jobs created by large professional firms are limited, jobs created by micro stores—although plentiful—are inferior, resulting in labor market duality within the private sector between the primary and secondary labor markets. Second, since Saudi workers are not willing to take up jobs in micro stores, the shortage of national labor can be largely considered as a disguised shortage, perpetuated by micro stores. Third, the private sector has expanded mostly by creating redundant micro stores, dependent almost solely on foreign workers and low capital intensity. This has created path dependence, which is far from job creation and economic

diversification. Therefore, it is argued that these factors have contributed, among other things, to disguised shortages of national labor and a slowing down of the transition to a more efficient economy. They have also contributed to labor market duality between the primary and secondary labor markets, resulting in occupational and geographical immobility of labor, and ultimately, unemployment persistence.

- The impacts of the sponsorship system can be summarized as follows. First, the sponsorship system leads to the problems of adverse selection and moral hazard resulting from asymmetric information among stakeholders. Second, the sponsorship system impedes the free market mechanism by increasing labor frictions and inflexibility, which exacerbates the issue of labor immobility. Third, the sponsorship system exacerbates the over-dependence on low-skilled foreign workers, creating path dependence. Fourth, the sponsorship system, combined with the excessive number of micro stores, contributes to commercial concealment, preventing or slowing down the integration of national labor into the private sector.
- Over the course of 50 years, the government has made efforts to implement both hard and soft policies to correct the labor market; however, these policies seem ineffective. The private sector, for its part, has exhibited inability or apathy, failing to take the initiative and create decent jobs, despite endless support from the government.
- Several corrective measures can be suggested to reverse the impacts of market-distorting factors; however, two have been singled out given the urgency of the situation: (i) replacing the sponsorship system by a central recruitment agency and (ii) re-zoning or restricting micro stores within self-contained centers as well as creating consortiums between the government and private investors to establish TDCs for developing certain industries and product markets while recruiting and training national labor.

4. Methodology and Data Analysis

4.1 Introduction

In economics, research is typically conducted using the normative or positivist methodologies. The former is used when researchers examine “what should be” questions in problem-solving and subject matter economic research while the latter is used when researchers analyze “what is” questions (Ehrenberg & Smith, 2012; Ethridge, 2004). According to Angrist and Krueger (1999), about 80% of papers published in labor economics contain some empirical analysis, and two-thirds of these use micro data. The authors state that in the 1970s, the use of micro data became more popular than time series data, but by the mid-1990s the use of micro data was used about ten times more than time series data. Angrist and Krueger (1999) distinguish between two different, yet complementary, methods of empirical research in labor economics, namely, descriptive analysis and causal inference. Descriptive analysis lays out facts about the labor market that need to be addressed by theoretical reasoning. The authors emphasize the importance of descriptive analysis as an essential step that must come before theorizing. This is because it provides facts that are used to build on theories. The other type of empirical research is by causal inference; this aims to identify the effects of specific policy interventions or estimate behavioral relationships.

This study used descriptive analysis in Chapter 2 to establish facts about the Saudi labor market. Then in Chapter 3, a theoretical framework was proposed about the impact of micro stores and other market-distorting factors. For empirical analysis, the study uses the nested CES production function to estimate the elasticity of substitution among three production inputs: (i) Saudi labor, (ii) non-Saudi labor, and (iii) capital in different types of establishment sizes—micro and big establishments. To the best of the researcher’s knowledge, this is the first attempt, in the context of the Saudi labor market, in which the

nested CES production function has been estimated. Therefore, this has provided new insights and a deeper understanding of the Saudi labor market.

The three market-distorting factors discussed in Chapter 3 have played a crucial role in different aspects of the economy. However, because of the limited scope of this study, the distorting impacts of those factors on the labor market cannot all be tested empirically at once. The crowding-out effect of the public sector is prominent and can be examined—for example, by estimating the effect of government borrowing from the domestic money market on the private investment or estimating the multiplier of government expenditure (Al-Asmari, 2008; Wilson et al., 2012). Although the crowding-out effect of the public sector on the labor market is discussed and illustrated graphically in this study, it is not tested empirically for three reasons. First, the duality between the public and the private sectors, stemming from the significant differences in the working conditions between the two sectors, is a well-known fact that has already been established in the literature (Bin Obaid, 2003; International Monetary Fund, 1997; Ramady, 2010). Second, empirical testing of the crowding-out effect is beyond the scope of this study since it would require an econometric model that is different than the one used in this research. Finally, it appears that the government is aware of the limited capacity of the public sector as the first-and-last-resort employer of Saudi labor because of the recurring call to the private sector to increase Saudi employment.

Thus, the study focuses on other factors—the micro stores effects and the sponsorship system effects. The former refers to the number of micro stores compared to big establishments while latter is related to the number of foreign workers. The study posits that there is a dual labor market, with each labor market having different characteristics that critically affect the elasticity of substitution between production inputs. This has profound implications on Saudi employment. To estimate the elasticity of substitution and examine whether the elasticity of substitution differs in a dual labor market setting, an estimation of the production function for each labor market would be a more plausible method.

Hence, the study focuses on estimating different types of elasticity of substitution using two-level nested CES production functions, as discussed below.

4.2 Econometric Models and Research Hypotheses

Three main economic models are used to estimate the production function, namely, the Cobb-Douglas (Cobb & Douglas, 1928), the Leontief (Leontief, 1986), and the CES (Arrow et al., 1961) production functions. The CES functions have several advantages compared to the other two models. Koesler and Schymura (2015) mention three reasons for the popularity of the CES functions in recent studies. First, CES functions have mathematical features allowing for easier analytical treatment compared to restricted functions, such as translog functions. Second, CES functions are flexible in characterizing different economic behavior. Finally, CES functions can be easily calibrated to benchmark values. The CES function is also considered a general model since it includes the other two models as special cases. For instance, the CES production function can take the following form:

$$Y = \gamma \left[\delta x_1^{-\rho} + (1-\delta) x_2^{-\rho}\right]^{-\nu/\rho}, (4.1)$$

where Y is the output quantity, x_1, x_2 are the input quantities, and $\gamma, \delta, \nu, \text{and } \rho$ are the parameters of the model. Parameter $\gamma \in [0, \infty)$ determines the total factor productivity, $\delta \in [0,1]$ determines the optimal distribution of the inputs, $\nu \in [0, \infty)$ determines the elasticity of scale, and $\rho \in [-1,0) \cup (0, \infty)$ determines the CES: $\sigma = (1+\rho)^{-1}$. The CES function includes the three special cases as follows:

1. If $\rho \to 0$, then $\sigma \to 1$; hence the CES turns to the Cobb-Douglas production function.
2. If $\rho \to \infty$, then $\sigma \to 0$; hence the CES turns to the Leontief production function with fixed factor proportions.

3. If $\rho \to -1$, then $\sigma \to \infty$; hence the CES turns to a linear production function with perfect factor substitution, provided that the function is homogeneous of degree one (i.e., $\nu = 1$).

Of course, the CES function is not restricted to only two inputs; the general n-input CES function is as follows:

$$y = \gamma \left(\sum_{i=1}^{n} \delta_i \, x_i^{-\rho} \right)^{-\nu/\rho} \quad with \sum_{i=1}^{n} \delta_i = 1, (4.2)$$

However, one disadvantage of the n-input CES function is that the elasticity of substitution is exactly the same for every pair of inputs, making it less useful for empirical applications (Henningsen & Henningsen, 2011; Koesler & Schymura, 2015). Consequently, Sato (1967) developed a two-level (nested) CES function where one can construct a function with more than two inputs and still able to obtain different constant elasticities of substitutions between different pairs of inputs. The idea of the nested CES function is that the function can have two or more levels of CES functions by assuming one or more inputs as an output of other CES function. For example, by following the notation for parameters used in Henningsen and Henningsen (2011), a three-input nested CES production function can be written as follows:

$$Y = \gamma \left[\delta \underbrace{\left(\delta_1 x_1^{-\rho_1} + (1 - \delta_1) \, x_2^{-\rho_1} \right)}_{lower \; level \; CES}^{-\rho/\rho_1} + (1 - \delta) x_3^{-\rho} \right]^{-\nu/\rho}, (4.3)$$

$$\sigma_{x_1, x_2} = \frac{1}{1 + \rho_1}, (4.4)$$

$$\sigma_{(x_1 x_2), x_3} = \frac{1}{1 + \rho}, (4.5)$$

where Y is the output quantity, and x_1, x_2, x_2 are the input quantities, but here $x_1, \text{and } x_2$ are nested in a lower CES function. According to Henningsen and Henningsen (2011), the lower level CES produces the Hicks-McFadden (direct)

elasticity between x_1 and x_2 (equation 4.4) and the upper level produces the Allen-Uzawa (partial) elasticity between the composite x_1 and x_2 as a group and x_3 (equation 4.5). The authors distinguish between the two types of the elasticities of substitution as follows:

> While the Hicks-McFadden elasticity of substitution (also known as direct elasticity of substitution) describes the input substitutability of two inputs i and j along an isoquant given that all other inputs are constant, the Allen-Uzawa elasticity of substitution (also known as Allen partial elasticity of substitution) and the Morishima elasticity of substitution describe the input substitutability of two inputs when all other input quantities are allowed to adjust. (Henningsen & Henningsen, 2011, p. 3)

Along the same lines, Anderson and Moroney (1994) elaborate on the overall substitution by differentiating between “pure intraprocess substitution” and “interprocess substitution.” The intraprocess substitution is defined by the Hicks-McFadden direct elasticity of substitution whereas the combined intraprocess and interprocess adjustment is measured by the Allen-Uzawa partial elasticity of substitution. Further, Anderson and Moroney (1994) link the two types of elasticity to the gross and net elasticities of substitution. They state that intraprocess (Hicks-McFadden) elasticity of substitution corresponds to gross elasticity because inputs within a weakly separable subprocess adjust to a change in relative input prices when the subprocess output is held constant and all inputs outside the subprocess are fixed. In contrast, interprocess (Allen-Uzawa) elasticity of substitution corresponds to the net elasticity of substitution because all inputs within and outside the specific process are adjusted to the cost-minimizing level. Anderson and Moroney (1994) assert that the two elasticities are identical if, and only if, the nested technologies are of the Cobb-Douglas form.

4.2.1 Research Model

A three-input nested CES production function is used to estimate the direct elasticity between Saudi labor and capital, and the partial elasticity between composite inputs (Saudi labor and capital) and foreign labor in both micro and big establishments. Assuming constant returns to scale and the Hicks-neutral technological change, the three-input nested CES production function for a labor market can be expressed as follows:

$$Y_{i,t} = \gamma\, e^{\lambda t}\left[\delta\left(\delta_1 L_{i,t}^{-\rho_1} + (1-\delta_1)\, K_{i,t}^{-\rho_1}\right)^{-\rho/\rho_1} + (1-\delta)F_{i,t}^{-\rho}\right]^{-1/\rho},$$

$$for\ i^{th}\ industry, and\ year\ t \quad (4.6)$$

where Y is output measured by the value added in Saudi Riyals and adjusted for inflation using 2013 as the base year, L is the number of local/Saudi workers, K is the net capital formation in an industry in Saudi Riyals and adjusted for inflation (2013=100), F is the number of foreign workers, and t is the time variable in years covering the period from 2010 to 2017 for 83 sub-industries. Parameter λ denotes a constant (Hicks-neutral) annual rate of technological change. The other parameters are defined as in equation (4.1). Equation (4.6) is estimated for each establishment size (large and micro), assuming they represent the two labor markets (the primary and secondary labor markets). This procedure is used to provide deeper insights into the direct and partial elasticity of substitution between Saudi and non-Saudi laborers in different labor markets.

Moreover, the nesting structure in equation (4.6) is modeled as $(LK)F$. This means that the Saudi workers and capital are combined in a lower CES function, and then the result is combined with foreign workers. Because the elasticities of substitution in nested CES functions are sensitive to the nesting structure, and because the nesting structure is theoretically arbitrary (Henningsen & Henningsen, 2011; Heun et al., 2017; Van der Werf, 2008), it is posited that the nesting structure as in (4.6) is more suitable for the purposes of this study for several reasons. First, the composite inputs play an important role in determining

the type of foreign workers, who are assumed to be skilled in the primary labor market and unskilled in the secondary labor market. Second, it is expected that the decision on the size of capital will be jointly determined with the decision of Saudi or foreign workers, but Saudi workers would more suitable because of the localization policies. Third, the nesting procedure should be done so that inputs in the lower level are complementary in nature (Khan, 1989). Because of the assumption that Saudi labor is relatively more skillful than foreign labor, and because skilled labor is considered to complement capital, nesting Saudi labor with capital would be more suitable than other combination structures. Finally, the partial elasticity of substitution measures the impact on one quantity of one price change rather than the connection between an input ratio and a price ratio measured by the direct elasticity of substitution (Stern, 2011). Hence, modeling foreign labor alone (i.e., not combined with other inputs) would enable us to predict the effect of the change in foreign labor costs in different labor markets resulting from the recent legislation of the expatriate levy.

4.2.2 Research Propositions and Hypotheses

Several studies testing the hypothesis of dual labor markets focus on whether wage differentials exist between the two labor markets.[18] However, in this study, a different approach was adopted for several reasons. First, the lack of disaggregated data does not allow us to test such hypotheses. Second, dividing the labor market into high- and low-wage labor markets to estimate wage differentials would result in the presence of wage differentials even if, in fact, there is no labor market segmentation. In other words, the estimation would be biased because of the truncated wage problem (Cain, 1976). This approach was avoided by estimating the direct and partial elasticities of substitution in different labor markets using the nested CES production function. Finally, one of the research purposes is to evaluate the replacement of non-Saudi workers with Saudi workers in different-sized establishments. Hence, the study focuses on

[18] For a comprehensive survey of the literature, see (Leontaridi, 1998).

two sets of establishment sizes—micro and large—that are presumed to capture positive and negative characteristics defining the two labor markets. Using firm size as a proxy for unobserved determinants could be a valid instrument, especially since it is widely used in the literature, for example, Oi (1990), Schmidt and Zimmermann (1991), and Krueger and Summers (1986). The basic premise is that those characteristics critically influence the elasticity of substitution between Saudi and non-Saudi workers. Hence, it is crucial that we pay close attention to the substitutability relationship between Saudi and non-Saudi labor in different labor markets since this holds critical implications for the analysis.

To elaborate, in the primary labor market, the relationship between Saudi and non-Saudi labor can be substitutable or complementary according to the type of skills or occupations. For example, if Saudi and non-Saudi laborers have similar skills (such as middle-level jobs), the relationship would substitutable. Hence, replacing foreign workers would be possible because Saudi workers tend to be willing to work in such jobs. In contrast, if foreign workers have specific skills that Saudi workers do not possess (such as upper-end jobs) or if they hold certain jobs that Saudi workers do not usually accept (such as lower-end jobs), the relationship would be complementary.

By contrast, in the secondary labor market, the relationship between Saudi and non-Saudi laborers would be mostly of a complementary nature because of the negative characteristics of this type of labor market. However, this relationship is complementary because the micro stores model resembles the passive income model rather than an entrepreneurial one, where Saudi workers play the role of employers (sponsors) and foreign workers play the role of employees (renters). In fact, if such a relationship exists, the relationship between Saudi and non-Saudi laborers might become a perfect complementary one as in the Leontief production function. Hence, the replacement mechanism would be expected to fail when the relationship between Saudi and non-Saudi laborers were a complementary relationship. Enforcing localization policies would be counterproductive and lead to severe ramifications. Therefore, the basis of the research is to estimate the employability of Saudi labor in different

labor markets, which depends on the elasticity of substitution considering different labor markets.

The research hypotheses are based on labor economic theory, where estimating the elasticity of substitution between different production inputs plays an important role in the analysis. In this study, estimating the elasticity of substitution between Saudi labor, capital, and non-Saudi labor is obtained by the nested CES production function. This function has been increasingly used in the literature since the 1990s, as can be seen in environmental economics literature examining the relationship between energy, capital, and labor (Chang, 1994; Dissou et al., 2015; Prywes, 1986; Van der Werf, 2008). It can also be seen in labor economics literature examining the relationships between skilled labor and capital (i.e., complementarity hypothesis between skilled labor and capital), between skilled and unskilled workers, or between national and migrant workers (Borjas, 2003, 2013; Card & Lemieux, 2001; Lewis & Peri, 2015). Borjas (2014) states that:

> The nested CES framework has two features that make it extremely useful for empirical analysis. First, it greatly reduces the dimensionality of the problem … Second, the nested CES framework is easily estimable using the national level census data. (p. 107)

As discussed in section 2.4.2.2, compelling evidence confirms that the elasticity of substitution between Saudi and non-Saudi labor is low, hence localization programs are considered ineffective. However, almost all those studies are descriptive or based on the inductive method. To the best of the researcher's knowledge, only one study used empirical evidence—that of Mashaal (2013). He estimated the (general) elasticity of substitution between Saudi and non-Saudi labor using the CES function and found that the elasticity of substitution was below unity at 0.457 and statically significant, indicating the complementarity relationship (i.e., low substitutability) between the two types of labor. This study provides empirical evidence examining the relationships between Saudi labor, capital, and non-Saudi labor while considering different labor markets. Hence, the research hypotheses are formulated to estimate the

direct and partial elasticities as well as the difference of these elasticities between the two labor markets. Presented below is a list of theoretical propositions and statements of hypotheses to be tested using equation (4.6).[19]

A. Micro Industries:

1) The study postulates that the direct elasticity of substitution between Saudi labor and capital in micro establishments will be zero (expect NOT to reject the null hypothesis):

$$H_0: \left(\sigma_{L,K}\right)_M = 0 \text{ (i.e., perfect complements)}$$

$$H_1: \left(\sigma_{L,K}\right)_M \neq 0 \text{ (i.e., not perfect complements)}$$

2) The study postulates that the partial elasticity of substitution between the composite inputs (Saudi labor and capital) and foreign labor in micro establishments will be zero (expect NOT to reject the null hypothesis):

$$H_0: \left(\sigma_{(LK),F}\right)_M = 0 \text{ (i.e., perfect complements)}$$

$$H_1: \left(\sigma_{(LK),F}\right)_M \neq 0 \text{ (i.e., not perfect complements)}$$

3) The study postulates that the production function in micro industries should not be specified as a Cobb-Douglas form (expect to reject the null hypothesis):

$$H_0: \left(\sigma_{L,K}\right)_M - \left(\sigma_{(LK),F}\right)_M = 0 \text{ (i.e., a Cobb-Douglas form)}$$

$$H_1: \left(\sigma_{L,K}\right)_M - \left(\sigma_{(LK),F}\right)_M \neq 0 \text{ (i.e., not a Cobb-Douglas form)}$$

B. Large Industries:

4) The study postulates that the direct elasticity of substitution between Saudi labor and capital in big establishments will be less than one (expect to reject the null hypothesis):

[19] Note: the subscripts B and M indicate the big and micro establishments, respectively.

$H_0: (\sigma_{L,K})_B - 1 \geq 0$ (i.e., a substitution relationship)

$H_1: (\sigma_{L,K})_B - 1 < 0$ (i.e., a complementary relationship)

5) The study postulates that the partial elasticity of substitution between the composite inputs (Saudi labor and capital) and foreign labor in big establishments will be bigger than one (expect to reject the null hypothesis):

$H_0: (\sigma_{(LK),F})_B - 1 \leq 0$ (i.e., a complementary relationship)

$H_1: (\sigma_{(LK),F})_B - 1 > 0$ (i.e., a substitution relationship)

6) The study postulates that the production function in big industries should not be specified as a Cobb-Douglas form (expect to reject the null hypothesis):

$H_0: (\sigma_{L,K})_{\boldsymbol{B}} - (\sigma_{(LK),F})_B = 0$ (Cobb-Douglas form)

$H_1: (\sigma_{L,K})_{\boldsymbol{B}} - (\sigma_{(LK),F})_B \neq 0$ (not Cobb-Douglas form)

C. Testing for Establishment Size on the Elasticities of Substitution in Micro and Large Industries:

7) The study postulates that the direct elasticity of substitution in micro establishments will be less than that in big establishments (expect to reject the null hypothesis):

$H_0: (\sigma_{L,K})_M - (\sigma_{L,K})_B \geq 0$ (i.e., the relationship between Saudi labor and capital is less complementary in micro industries than that in big industries)

$H_1: (\sigma_{L,K})_M - (\sigma_{L,K})_B < 0$ (i.e., the relationship between Saudi labor and capital is more complementary in micro industries than that in big industries)

8) The study postulates that the partial elasticity of substitution in micro establishments will be less than that in big establishments (expect to reject the null hypothesis):

$H_0: (\sigma_{LK,F})_M - (\sigma_{LK,F})_B \geq 0$ (i.e., the relationship between the composite inputs and foreign labor is less complementary in micro industries than that in big industries)

$H_1: (\sigma_{LK,F})_M - (\sigma_{LK,F})_B < 0$ (i.e., the relationship between the composite inputs and foreign labor is more complementary in micro industries than that in big industries)

Hypotheses 1 and 4 are formulated to test the direct elasticity of substitution between Saudi labor and capital in micro and big establishments, respectively. These two hypotheses are related to the complementarity hypothesis between skilled/Saudi labor and capital with respect to the two labor markets. Hypotheses 2 and 5 are formulated to test the partial elasticity of substitution between the composite inputs (Saudi labor and capital) and foreign labor in micro and big establishments, respectively. These two hypotheses are related to the substitutability hypothesis between the composite inputs (Saudi labor and capital) and non-Saudi labor. Hypotheses 3 and 6 are formulated to test the Cobb-Douglas form of the production function, as usually conducted in the literature, although they are not the main focus of this study. Finally, hypotheses 7 and 8 are formulated to test the difference of each elasticity in different labor markets to assess whether different labor markets have implications for Saudi employability. All the research hypotheses are tested statistically using a 5% significance level, as usually adopted in empirical economic studies.

4.3 Data Sources and Estimation Method

4.3.1 Data Sources

As one important contribution of this study, it uses the Establishments Economic Survey that is annually published by GASTAT and is based on the

Enumerating Establishments census. This contribution is significant because, to the best of the researcher's knowledge, until date, no study has used the Enumerating Establishments' census despite its paramount importance in enhancing the understanding of the Saudi labor market structure. This study uses the 2010 census for the descriptive analysis and uses the survey series (2010–2017) for the empirical analysis. The Enumerating Establishments was the first-ever census of establishments, published in 2010 by General Authority for Statistics (2010a). It covers the number of establishments, Saudi workers, non-Saudi workers, compensation, expenditure, value added and gross capital formation. All variables are disaggregated according to three categories of establishment sizes, which are classified according to the number of employees (less than 5, 5 to 19, and 20 or more) except for gross capital formation, which is reported in total (aggregated). The data are time series, consisting of 83 cross-sectional industries, listed according to ISIC4,[20] and eight years covering the period 2010–2017 (General Authority for Statistics, 2017a). Therefore, there are 83 sub-industrial sectors and eight years, yielding a total of 664 observations. After exploring the data, industry number 5 lacked enough data, causing some errors in the estimation. Hence, it was excluded from the analysis, yielding 656 total observations.

The study uses only two establishment sizes (less than 5 and 20 or more employees) for micro and big establishment sizes to distinguish between the secondary and primary labor markets, respectively. Although there seems to be no consensus on the definition of firm size (Shalit & Sankar, 1977), this classification may not be the ideal classification for the research objectives. In particular, using 20 or more employees to define a big establishment may not be an optimal definition for big establishments nor for testing the research hypotheses. However, because of the absence of data sources enabling us to freely choose the number of employees at this stage, it was decided to proceed and make use of the available data. Further, the Annual Statistics series

[20] See Appendix 3.

published by SAMA were used to calculate the real values of variables using the deflator index shown in Table 4.1.

Table 4.1: CPI and Deflator Index (2013 = 100)

Year	CPI	Deflator Index (DI)
2010	90.5	0.905
2011	93.9	0.939
2012	96.6	0.966
2013	100	1
2014	102.2	1.022
2015	103.5	1.035
2016	105.6	1.056
2017	104.7	1.047

Source: (Saudi Arabian Monetary Authority, 2017)

4.3.2 Estimation Method

In terms of estimation, there are two approaches for estimating the substitution parameters, a technical (direct) approach and an economic (indirect) approach, both of which have their own drawbacks (Mizon, 1977). While the former is based on direct estimation by nonlinear least squares, the latter is based on the first-order conditions derived from either the profit-maximization or cost-minimization objective functions. The economic approach requires the availability of all input prices and the absence of market distortion (Henningsen et al., 2019). Although both approaches have their own drawbacks, Mizon (1977) maintains that “direct estimation, by non-linear least squares when necessary, is feasible and often preferable to indirect estimation” (p. 1240). Moreover, Kumar and Gapinski (1974a, 1974b) examined the small-sample properties of the nonlinear least-squares estimators of the CES production function parameters. By using the Monte Carlo technique, they drew 150 sets of 20 observations and

found that the bias for the estimators of all parameters, except for σ, was always small, regardless of whether assuming multiplicative or additive error term. They concluded by stating:

> Nonlinear least squares appears to be an important tool for estimating the parameters of the CES production function as it seems to exhibit a tendency for providing accurate estimates, except for σ. This finding should prove useful to the econometrician who plays in the world of nonlinear techniques. (Kumar & Gapinski, 1974b, p. 567)

The imprecise estimation of σ by using the nonlinear least squares is attributed to the tendency to flatness of the objective functions around the minimum (Henningsen & Henningsen, 2011; Kumar & Gapinski, 1974b). However, Henningsen and Henningsen (2011) suggest that "this problem can be alleviated by performing a grid search" (p. 27), a recommendation that was followed in this study.

Therefore, in our case, the model was estimated using the nonlinear least squares procedure for several reasons. First, the conditions of the economic approach conflict with the fundamental premise of this study, that three distorting factors exist in the Saudi labor market. Second, there was a lack of disaggregated data of wage levels and prices of different inputs, enabling the use of first-order conditions. Third, despite the drawbacks of this procedure, it would provide—as the first study in this regard—initial estimates for the elasticity of substitution between the three production inputs. Further studies could build on these findings or compare their results with them. Fourth, the nonlinear least squares estimation is widely used in the literature (Henningsen & Henningsen, 2011, 2012; Heun et al., 2017; Koesler & Schymura, 2015; KUCUK et al., 2018; Shen & Whalley, 2013). Finally, because of the tendency of flatness of the objective functions around the minimum when using the nonlinear least squares, Henningsen et al. (2019) suggest that these regressions "require (i) large data sets and (ii) large independent variations in the explanatory variables … in order to obtain reliable estimates" (p. 25). These two requirements are, to a greater extent, met in this study because of the following three reasons. (i) The study

uses panel data sets which yield a total of 656 observations (for each industry). This makes the number of observations higher than those studies examined in their study (Henningsen et al., 2019). (ii) In terms of the multicollinearity problem, the results of the variance inflation factor test indicate that there is no multicollinearity between different explanatory variables as explained in the next section. (iii) Finally, the research implements different optimization algorithm methods with a grid search for a robustness check, and for mitigating the flatness problem around the minimum.

Moving to the model specification, equation (4.6) is estimated by adding an error (stochastic) term to the right-hand side of the equation, and then finding the best fit of the equation that minimizes the sum of squared residuals. That is:

$$\min_{\gamma,\lambda,\delta,\delta_1,\rho_1,\rho} \sum_t \left(Y_t - \gamma\, e^{\lambda t} \left[\delta \left(\delta_1 L_t^{-\rho_1} + (1-\delta_1)\, K_t^{-\rho_1} \right)^{-\rho/\rho_1} + (1-\delta) F_t^{-\rho} \right]^{-1/\rho} \right)^2, \quad (4.7)$$

Because objective function (4.7) is nonlinear in parameters, it requires a nonlinear optimization algorithm with the aid of statistical software. Thus, the statistical software R Project (Version 1.3.1093, 2020-09-17) was used as well as the "micEconCES" package developed by Henningsen and Henningsen (2011) to estimate the CES functions. The micEconCES package addresses some issues of estimating nonlinear functions such as (i) finding the global minimum of the objective function, and (ii) estimating the parameters that are consistent with economic theory (Henningsen et al., 2019). To increase the likelihood of finding the set of parameters that result in a global minimum of the objective function, three different optimization algorithm methods were used. In particular, the first two were unrestricted and restricted gradient-based optimization algorithms. The unrestricted algorithm was the Levenberg-Marquardt algorithm (LMA) while the restricted one was the PORT routines algorithm. A third gradient-free global optimization algorithm was the Differential Evolution optimization algorithm (DEA), which was used for a robustness check. The boundaries of the PORT and DEA were within an economically meaningful region. By adopting the estimation procedure in Henningsen et al. (2019) with

little changes, the estimation procedure can be summarized in the following steps:

1) Estimating the nested CES function, using the micEconCES package in R, by running a two-dimensional grid search with 11 different values between -1 and 50 for each of ρ_1 and ρ, i.e., for elasticities of substitution between 0.02 and infinity. For each of the $11^2 = 121$ estimations, the values of ρ_1 and ρ are fixed at values at the respective grid point while the other parameters, if present, are estimated using the default starting values ($\lambda = 0.015, \delta_1 = \delta = 0.5, \nu = 1$ and γ is set to a value so that the mean of the error term is zero (for more details, see Henningsen and Henningsen (2015)). Of these 121 estimations, the estimation that results in the smallest sum of squared residuals is chosen.

2) Running a two-dimensional grid search using the rho's parameters obtained in approach 1 as center values of a new grid range with a width equal to 10 (rho's ∓ 10 range) and grid step smaller than 1. For each of the $21^2 = 441$ estimations, the values of ρ_1 and ρ are fixed at values at the respective grid point, while the other parameters ($\gamma, \lambda, \delta_1, \delta, \nu$), if present, are estimated using the default starting values.

3) Estimating the nested CES function using the parameters obtained in approach 2 as starting values.

4) Estimating the nested CES function using the default starting values used by the micEconCES package.

5) Choosing between a model with a grid search starting point (approach 3) and with the default starting point (approach 4) based on the principle of minimization of the sum of squared residuals.

Therefore, the scaled (i.e., log data) and unscaled data to fit the model yields six fitted models for each industry. Of these models, the best-fitted model is the one that produces the least sum of squared residuals (i.e., the lowest residual standard error and the highest R-squared), provided that its parameters have

meaningful economic interpretations. The goodness of fit, convergence, and compliance to economic theory (i.e., parameters within economic bounds) are three criteria that can be used for selecting the appropriate model in the nonlinear estimation methods (Feng & Zhang, 2018).

4.4 Conclusion

- The main ideas discussed in this chapter can be summarized as follows:The study focuses on two sets of establishment size, micro and big, that are presumed to capture positive and negative characteristics defining the two labor markets. The basic premise is that these characteristics critically influence the elasticity of substitution between Saudi and non-Saudi workers.
- It is crucial to pay close attention to the substitutability relationship between Saudi and non-Saudi labor in different labor markets since this has critical implications for the analysis. In the primary labor market, the relationship between Saudi and non-Saudi labor can be substitutional or complementary, depending to the type of skills or occupations. For example, if Saudi and non-Saudi laborers have similar skills (such as middle-level jobs), the relationship would be substitutional. Hence, replacing foreign workers would be possible because Saudi workers are willing to work in such jobs. However, if foreign workers have specific skills that Saudi workers do not possess (such as upper-end jobs) or if they hold certain jobs that Saudi workers do not usually accept (such as lower-end jobs), the relationship would be complementary.
- In contrast, in the secondary labor market, the relationship between Saudi and non-Saudi laborers is mostly complementary because of the negative characteristics in this type of labor market. It is also considered a complementary relationship because the micro stores model resembles the passive income model, rather than an entrepreneurial one, in which Saudi workers play the role of employers (sponsors) and foreign workers play the role of employees (renters).
- The study uses of the Establishments Economic Survey that is annually published by GASTAT and based on the Enumerating Establishments. The Enumerating Establishments was the first-ever census of establishments, published in 2010 by GASTAT. The study uses only two establishment sizes (5< and 20+) for micro and big establishments to distinguish between the secondary and primary labor markets.

The study also uses a three-input nested CES production function to estimate the direct elasticity between Saudi labor and capital, and the partial elasticity between composite inputs (Saudi labor and capital) and foreign labor in both micro and big establishments.

5. Results and Implications of the Study

5.1 Introduction

This chapter is divided into three sections: (i) descriptive statistics and the variance inflation factor, (ii) parameter estimates and the elasticity of substitution including the hypothesis testing, and (iii) the implications of the study. Apart of reporting empirical results and estimation, this chapter aims to discuss possible implications of the study on the Saudi labor market.

5.2 Descriptive Statistics and Variance Inflation Factor

The descriptive statistics for the different variables of the production function in micro and big industries are presented in Table 5.1. Those statistics were measured using data across 83 micro and big industries over the period 2010 to 2017. These descriptive statistics reveal three key features. First, the difference in output between the two industries can be seen in the median estimates because the added value in big industries is about 16 times more than that of micro industries. This is despite the fact that the firm size distributions for micro and big industries are 83% and 3%, respectively. Second, in terms of the number of workers, big industries, by definition, have more labor (Saudi and non-Saudi). However, in both micro and big industries, these establishments recruit non-Saudi labor more than Saudi labor. Finally, the net capital formation is aggregated; hence, we cannot comment on each industry size separately. However, the median value of the net capital formation is about SR645,000 only, which may reveal the extent of low capital investment in the private sector. This may support the research argument that low capital investment is one of the prominent features of the private sector.

Table 5.1: Descriptive Statistics of Different Variables Using Data Across 83 Micro and Big Industries Over the Period 2010 to 2017

	Mean	**Std. Dev.**	**Min.**	**1st Qu.**	**Median**	**3rd Qu.**	**Max.**
M_Y	1587223	5590932.6	-5381	16709	85302	556806	47598780
B_Y	18467405	106249133	-32279	278110	1355504	7661949	1.262E+09
M_L	4306	15575.96	3	57	320	2074	157102
B_L	9272	13927.73	2	868	3028	11776	73675
M_F	15189.9	54792.03	7	184.8	693	4231	507661
B_F	24194	54522.51	18	2557	6500	22838	521726
K	3428646	7919953.8	0	114575	645275	2940302	60310157

Note: The prefixes M and B indicate micro and big industries, respectively, Y is the output measured by the value added in Saudi Riyals and adjusted for inflation (2013=100), L is the number of national laborers, F is the number of non-national laborers, and finally, K is the net capital formation in Saudi Riyals and adjusted for inflation (2013=100).

Table 5.2 and Table 5.3 represent the variance–covariance (symmetric) matrices in micro and big industries, respectively.

Table 5.2: Variance-Covariance Matrix for the Variables of Micro Industries

	M_Y	M_L	K	M_F
M_Y	3.125853e+13	78080912348	6.942983e+12	283363888414
M_L		242610671	1.690024e+10	804609348
K			6.272567e+13	62467452858
M_F				3002166973

Note: The diagonal elements show a high variation in each variable

Table 5.3: Variance-Covariance Matrix for Variables of Big Industries

	M_Y	M_L	K	M_F
M_Y	1.128888e+16	568855864129	5.973258e+14	112134518616
M_L		193981586	7.383999e+10	401096310
K			6.272567e+13	87589267223
M_F				2972703781

Note: The diagonal elements show a high variation in each variable

The diagonal elements in the matrices contain the variances of each variable while the off-diagonal elements contain the covariance of each pair of variables. The purpose of reporting these matrices is to assess the variances of each variable.[21] Hence, the diagonal elements reveal that there was high variation in each explanatory variable in both industries, enabling better identification of the effects of those variables.

Table 5.4 and Table 5.5 represent the correlation matrices in micro and big industries, respectively. The correlation matrix is a standardized version of the variance–covariance matrix for a better examination of the potential existence of a multicollinearity problem. The off-diagonal elements in the correlation matrix measure the direction and strength of the linear association of each pair of variables. Hence, by examining the two matrices, the only potential problem of

[21] We can also obtain the standard deviation of each variable by taking the square root of each variance variable (standard deviations are reported in Table 5.1).

multicollinearity appears to be between the local labor (M_L) and foreign labor (M_F) variables.

Table 5.4: Correlation Matrix for the Variables of Micro Industries

	M_Y	M_L	K	M_F
M_Y	1.000	0.897	0.157	0.925
M_L		1.000	0.137	**0.943**
K			1.000	0.144
M_F				1.000

Note: There may be multicollinearity between M_L and M_F variables (element indicated in bold)

Table 5.5: Correlation Matrix for the Variables of Big Industries

	B_Y	B_L	K	B_F
B_Y	1.000	0.384	0.710	0.019
B_L		1.000	0.670	0.528
K			1.000	0.203
B_F				1.000

Note: There seems to be no indication of multicollinearity

However, examining the multicollinearity problem through the correlation matrix could be considered imprecise. Hence, a more formal examination testing for multicollinearity is to use the variance inflation factor (VIF). The VIF "measures how much the variances of the estimated regression coefficients are inflated as compared to when the predictor variables are not linearly related" (Kutner et al., 2005, p. 408). The cut-off value by which we might conclude that there is a multicollinearity problem is when $VIF = 10$ (Kutner et al., 2005; Wooldridge, 2013). The results of VIF between the explanatory variables are shown in Table 5.6. Because all VIF values are under 10, it can be concluded that there is no multicollinearity among the explanatory variables in each industry. However, we should interpret the estimated coefficients in micro industries with caution because there is a high degree of linear association between the local labor (M_L) and foreign labor (M_F) variables.

Table 5.6: Results of Variance Inflation Factor Among Explanatory Variables in Micro and Big Industries

Micro Industries	M_L	K	M_F
	8.996	1.021	**9.014**
Big Industries	B_L	K	B_F
	2.556	1.922	1.471

Note: There is a high degree of linear association in the local labor (M_L) and foreign labor (M_F) variables with other explanatory variables (elements indicated in bold)

5.3 Parameter Estimates and the Elasticity of Substitution

Table 5.7 and Table 5.8 represent the overall results of the three optimization algorithmic methods, namely, the LMA, the DEA, and the PORT routines algorithm. The first three rows show results from unscaled (raw) data, the last three rows show results from scaled (natural logarithm) data while the rows in bold font indicate the best-fitted models.

Table 5.7: Results from the Fitted Models of the Micro Industry

	γ	λ	δ_1	δ	ρ_1	ρ	c	RSS	R^2
LMA grid	119.070**	0.057**	0.994**	0.465**	- 1.000**	0.800*	1	2.48E+15	0.879
DEA	**117.115****	**0.056****	**0.985****	**0.509****	**- 0.774****	**0.820***	**0**	**2.52E+15**	**0.877**
PORT	100.158**	0.058**	0.953**	0.527**	- 0.562*	0.733*	0	2.55E+15	0.876
LMA log grid	1.302**	0.008**	0.010	0.267	1.700*	- 1.000*	1	1506.068	0.514
DEA log	1.312**	0.007**	0.347	0.474	- 1.000	- 1.000	0	1526.316	0.507
PORT log	1.309**	0.007**	0.348	0.479	- 1.000	- 1.000	1	1504.858	0.514

Note: Row in bold font indicates the best-fitted model in terms of the least sum squared of residuals and the parameters within economic bounds. For the last three methods, scaled (natural log) data were used. */** indicates that the coefficient is significantly different from 0 at the 5% and 1% levels, respectively.

Table 5.8: Results from the Fitted Models of the Big Industry

	γ	λ	δ_1	δ	ρ_1	ρ	c	RSS	R^2
LMA grid	**1.433****	**0.201****	**0.000**	**0.990****	**4.400**	**0.800****	**1**	**1.04E+18**	**0.860**
DEA	16.732	- 0.035	0.387	0.979	- 0.954	- 0.987	0	3.75E+18	0.492
PORT	9.872**	- 0.037	0.000	1.000**	1.200	0.100	0	3.75E+18	0.493
LMA log grid	1.567**	0.004*	0.996**	0.938**	-12.196**	- 6.420	1	1393.5040	0.559
DEA log	1.400**	0.004*	0.493**	0.657**	- 1.000	- 1.000	0	1991.844	0.370
PORT log	1.386**	0.005*	0.473**	0.666**	- 1.000	- 1.000	0	1963.600	0.379

Note: Row in bold font indicates to the best-fitted model in terms of the least sum squared of residuals and the parameters within economic bounds. For the last three methods, scaled (natural log) data were used. */** indicates that the coefficient is significantly different from 0 at the 5% and 1% levels, respectively.

Table 5.7 represents the estimation of micro industries. The first three methods show robust estimates because there is no change in the estimated coefficients in terms of their signs and statistical significance. There is also no significant change in terms of magnitudes of the estimated coefficients except perhaps for $\gamma \text{ and } \rho_1$. All three models have high and stable values of R^2. The estimated coefficient of the time trend (λ) indicates that for eight years (2010–2017), the micro industries grew annually on average at 5%. The estimated coefficient of share distribution of local labor (δ_1) indicates low capital usage in micro industries, which may support the research postulation, as mentioned earlier.

In contrast, the last three methods (with log data) show robust estimates only for gammas and lambdas because their signs, statistical significance, and magnitudes do not change significantly. The other estimated coefficients changed in terms of their magnitudes, signs, or statistical significance. All three models had R-squared values of around (0.5). Optimal choice, by R-squared, is the first one model using the LMA method with a grid search (LMA grid), but that model shows $\rho_1 = -1$; hence, the direct elasticity of substitution $\left(\sigma_{L,K}\right)_M = \infty$. In this case, it is impossible to calculate the standard errors. Consequently, no hypothesis can be tested with that model. The next best-fitted model, by R-squared, is the second, with the DEA method, with raw data and $R^2 = 0.877$. Thus, the DEA model was chosen as the “winning” model (i.e. the best-fit models) for micro industries, considering the goodness of fit and the economic range of the parameters.

Table 5.8 represents the estimation of big industries. The first three methods show non-robust estimates for all parameters because there were some changes in their signs, statistical significance, or magnitudes. The first method (LMA grid) yielded a higher R-squared value than the other two models (DEA and PORT), and it was the only model that converged. In contrast, the last three methods yielded robust estimates for $\gamma, \lambda, \delta_1, and\ \delta$ in terms of their signs and statistical significance, not magnitudes. The LMA log grid yielded estimates for $\rho\ and\ \rho_1$

that were far from the economic bound. Overall, these last three methods using log data produced poor R-squared values; hence, the estimates are likely to be imprecise. Therefore, the first method using raw data (the LMA grid) was chosen as the "winning" model for big industries, considering the goodness of fit and the economic range of the parameters. Having discussed the various optimization algorithmic methods, we move to discuss the results of the best-fit model for each micro and big industry.

5.3.1 The Best-Fit Model in Micro Industries

By using the "cesEst" function, the three-input nested CES production function was estimated using micro-industry data, while the definition of variables is as described in the research model, equation (4.6). The statistical summary from estimating the model using the DEA method is presented in Table 5.9.

Table 5.9: Statistical Summary of Estimating the nested CES Production Function in Micro Industries Using the DEA Method

	Estimate	Std. Error	t value	Pr $(>\|t\|)$
γ	117.115	16.317	7.177	0.000***
λ	0.056	0.012	4.819	0.000***
δ_1	0.985	0.022	44.575	0.000***
δ	0.509	0.074	6.885	0.000***
ρ_1	-0.775	0.300	-2.577	0.010**
ρ	0.820	0.364	2.250	0.024*
Residual standard error: 1958654				
Multiple R-squared: 0.8770837				
Elasticities of Substitution				
	Estimate	Std. Error	t value	Pr $(>\|t\|)$
σ_{HM}	4.420	5.866	0.754	0.451
σ_{AU}	0.550	0.110	4.995	0.000***

Note: Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
σ_{HM} = Hicks-McFadden (direct) elasticity of substitution
σ_{AU} = Allen-Uzawa (partial) elasticity of substitution

At first glance, it appears that all estimated coefficients are statistically significant except for the direct elasticity of substitution. The estimation of the efficiency parameter (γ) is relatively high, which may indicate efficient usage of the production inputs. Whereas the estimation of the rate of technological change (λ) is relatively low, which indicates that growth, in micro industries, may be explained well by endogenous factors of production. Both estimated parameters may suggest that the current conditions of micro stores may not require technological advancement since they meet their business purposes without the need of incurring extra costs. However, micro stores might not be able to sustain any requirements (e.g., by the government) that result in an increase in their production costs such as upgrading their production technology and localization policies.

In terms of elasticities of substitution, the estimated direct elasticity of substitution is high ($\hat{\sigma}_{S,K} = 4.420$), indicating relative ease of substitution. The high degree of substitution between Saudi labor and capital in micro industries is unexpected because Saudi labor tends to be the financier in micro stores; hence, the two should be complementary. However, the statistical test is in favor of the perfectly complementary relationship between local labor and capital in micro industries, as discussed in the next section. Another explanation for this high degree of substitution is that Saudi labor can be considered unskilled labor in that industry. This is because it is known in the economic literature that capital tends to come as a complement input with skilled labor, while it comes as a substitute input with unskilled ones. However, because the estimated value is statistically insignificant, we cannot draw a conclusion with high certainty. In contrast, the estimated partial elasticity of substitution is lower than unity ($\hat{\sigma}_{(SK),F} = 0.550$) and statistically significant, indicating the complementarity relationship between composite inputs (Saudi labor and capital) and foreign labor.

5.3.2 The Best-Fit Model in Big Industries

By using the "cesEst" function, the three-input nested CES production function was estimated using big-industry data, while the definition of variables is as described in the research model, equation (4.6). The statistical summary from estimating the model using the LMA method with a grid search is presented in Table 5.10.

Table 5.10: Statistical Summary of Estimating the Nested CES Production Function in Big Industries Using the LMA Method

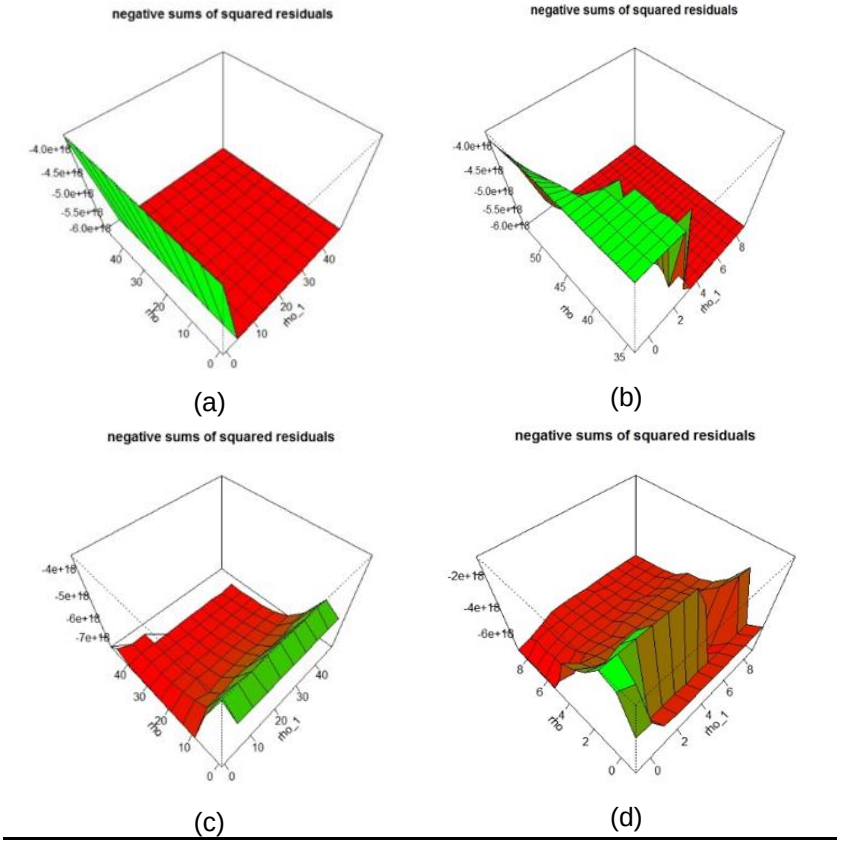
	Estimate	Std. Error	t value	Pr $(>\|t\|)$
γ	1.432	0.231	6.203	0.000***
λ	0.201	0.031	6.553	0.000***
δ_1	0.000	6.359e-13	0.026	0.979
δ	1.000	5.492e-04	1822.319	0.000 ***
ρ_1	4.400	5.473	0.804	0.421
ρ	0.800	0.079	10.139	0.000 ***
Residual standard error: 1958654				
Multiple R-squared: 0.8770837				
Elasticities of Substitution				
	Estimate	Std. Error	t value	Pr $(>\|t\|)$
σ_{HM}	0.185	0.188	0.987	0.324
σ_{AU}	0.556	0.024	22.812	0.000 ***

Note: Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
σ_{HM} = Hicks-McFadden (direct) elasticity of substitution
σ_{AU} = Allen-Uzawa (partial) elasticity of substitution

Here, few estimated coefficients are not statistically significant, including the direct elasticity of substitution. The estimation of the efficiency parameter (γ) is higher than one, which may still indicate efficient usage of the production inputs. Whereas the estimation of the rate of technological change (λ) is relatively high, which indicates that growth, in big industries, may not be explained well by endogenous factors of production. Hence, policymakers may need to encourage more investment in technology and innovation to accelerate growth. In terms of elasticities of substitution, the estimated direct elasticity of substitution is below

unity ($\hat{\sigma}_{S,K} = 0.185$), indicating a low degree of substitution between the Saudi labor and capital. However, because the estimated value is statistically insignificant, we cannot draw a conclusion with certainty. In contrast, the estimated partial elasticity of substitution is below unity ($\hat{\sigma}_{(SK),F} = 0.556$) and statistically significant. It also indicates a low degree of substitution between composite inputs (Saudi labor and capital) and foreign labor. Unlike the direct elasticity in both industries, the partial elasticity of substitution in both industries is statistically significant. All elasticities of substitution, except for the direct elasticity of substitution in micro industries, are lower than unity.

Figure 5.1: Goodness of Fit for Different Values of ρ_1 and ρ Using a Two-Step Grid Search



Note: (a) and (b) represent the two-dimensional grid search for ρ_1 and ρ in micro industries while (c) and (d) represent the two-dimensional grid search for ρ_1 and ρ in big industries.

Figure 5.1 represents a two-step grid search for finding the best fit of the parameters. Because the objective function for estimating the CES function by nonlinear least squares tends to have a flat surface around the minimum, the

grid search can be used to mitigate such a problem. Consequently, a grid of values for the substitution parameters (ρ_i) is pre-selected while other parameters are estimated. Then, “[t]he estimates with the values of the substitution parameters that result in the smallest sum of squared residuals are chosen as the final estimation results” (Henningsen & Henningsen, 2011, p. 27).

5.3.3 Hypothesis Testing

Table 5.11 reports the estimated values of the two types of elasticity of substitution in both sizes of establishments (micro and big) and the difference of elasticity of substitution between the two establishment sizes along with their standard errors, confidence intervals, and p-values. All hypothesis tests have generally narrow confidence intervals except for the hypothesis tested in 1, 3, and 7. A narrow confidence interval indicates the precision of the estimates if we assume all assumptions of the statistical model are correct. For example, given a confidence coefficient of $1 - \alpha = .95$, the confidence interval means that in a repeated sampling of size N, approximately 95% of all intervals would include the true mean (Wackerly et al., 2008).

Table 5.11: Results of the Hypothesis Testing				
	Estimated Value	Std. Error	95% Confidence Interval	p-value
Micro Establishments				
1) Direct Elasticities of Substitution	4.420	5.866	- 7.098; 15.938	0.451
2) Partial Elasticities of Substitution	0.550	0.110	0.334; 0.766	0.000***
3) Cobb-Douglas	3.870	5.853	- 7.622; 15.363	0.509
Large Establishments				
4) Direct Elasticities of Substitution	0.185	0.188	- 0.183; 0.554	0.000***
5) Partial Elasticities of Substitution	0.556	0.024	0.508; 0.603	0.999
6) Cobb-Douglas	- 0.370	0.168	- 0.700; 0.040	0.028*
Establishment Size Differences				
7) Direct Elasticities of Substitution	4.235	5.869	- 7.2885; 15.759	0.765
8) Partial Elasticities of Substitution	- 0.006	0.113	- 0.2273; 0.215	0.479
Note: Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1				

Below, the null hypotheses are recalled, with comments on each one, assuming that all assumptions of the statistical model are correct and using a 5% significance level, as usually adopted in empirical economic studies.

1) $H_0: \left(\sigma_{L,K}\right)_M = 0$ (i.e., there is a perfect complementary relationship between local labor and capital in micro industries). The point estimate of the direct elasticity of substitution in micro industries is about 4.42 with a p-value of 0.45. Because the p-value is greater than 0.05, the null hypothesis cannot be rejected. That is, there is evidence in favor of the perfectly complementary relationship between local labor and capital in micro industries. However, the wide confidence interval reflects an imprecise estimation.

2) $H_0: \left(\sigma_{(LK),F}\right)_M = 0$ (i.e., there is a perfect complementary relationship between the composite inputs and foreign labor in micro industries). The

point estimate of the partial elasticity of substitution in micro industries is about 0.55 with a p-value of 0.00. Because the p-value is less than 0.05, the null hypothesis can be rejected. That is, the composite inputs and foreign labor are not perfect complements.

3) $H_0: \left(\sigma_{L,K}\right)_M - \left(\sigma_{(LK),F}\right)_M = 0$ (i.e., the production function in micro industries should be specified as a Cobb-Douglas form). The point estimate of the difference between the two types of elasticity of substitution is about 3.87 with a p-value of 0.51. Because p-value is greater than 0.05, the null hypothesis cannot be rejected. That is, the production function should be a Cobb-Douglas function. However, the wide confidence interval reflects an imprecise estimation.

4) $H_0: \left(\sigma_{L,K}\right)_B - 1 \geq 0$ (i.e., there is a substitution relationship between local labor and capital in big industries). The point estimate of the partial elasticity of substitution in micro industries is about 0.185 with a p-value of 0.00. Because the p-value is less than 0.05, the null hypothesis can be rejected. That is, there is evidence in favor of the complementary relationship between local labor and capital in big industries.

5) $H_0: \left(\sigma_{(LK),F}\right)_B - 1 \leq 0$ (i.e., there is a complementary relationship between the composite inputs and foreign labor in big industries). The point estimate of the partial elasticity of substitution in big industries is about 0.56 with a p-value of 0.999. Because the p-value is greater than 0.05, the null hypothesis cannot be rejected. That is, there is evidence in favor of the complementary relationship between the composite inputs and foreign labor in big industries.

6) $H_0: \left(\sigma_{L,K}\right)_{\boldsymbol{B}} - \left(\sigma_{(LK),F}\right)_B = 0$ (i.e., the production function in big industries should be specified as a Cobb-Douglas form). The point estimate of the difference between the two types of elasticity of substitution is about −0.37 with a p-value of 0.028. Because the p-value is less than 0.05, the null hypothesis can be rejected. That is, the production function in big industries should not be a Cobb-Douglas form.

7) $H_0: (\sigma_{L,K})_M - (\sigma_{L,K})_B \geq 0$ (i.e., the relationship between Saudi labor and capital in micro industries is less complementary than that in big industries). The point estimate of the difference between the two estimates of direct elasticity is about 4.24 with a p-value of 0.77. Because the p-value is greater than 0.05, the null hypothesis cannot be rejected. That is, the relationship between Saudi labor and capital in micro industries is less complementary than that in big industries.

8) $H_0: (\sigma_{LK,F})_M - (\sigma_{LK,F})_B \geq 0$ (i.e., the relationship between the composite inputs and foreign labor in micro industries is less complementary than that in big industries). The point estimate of the difference between the two estimates of partial elasticity is about −0.006 with a p-value of 0.48. Because the p-value is greater than 0.05, the null hypothesis cannot be rejected. That is, the relationship between the composite inputs and foreign labor in micro industries is less complementary than that in big industries.

In big industries, both the direct and partial elasticities of substitution have narrow estimate intervals, which may indicate the higher precision of their estimates. In micro industries, by contrast, the estimate for the partial elasticity of substitution might be more precisely estimated than the estimate of the direct elasticity of substitution because the latter has a considerably wider confidence interval. This might be the reason for the failure to reject the first, third, seventh, and eighth null hypotheses since they involve the value of the direct elasticity of micro industries.

Further, the research findings suggest that there might be no difference between micro and big industries in terms of substitution degree between the two groups, which is contradictory to the research hypotheses (in 7 and 8). One likely explanation for this is that the data used for big industries might not be the best representative of such industries. The study adopted the official classifications for micro and big industries (i.e., fewer than 5 employees for micro industries and

20 employees or more for big industries). Although there seems to be no consensus on the definition of firm size (Shalit & Sankar, 1977), the classification of big firms may not be the perfect classification for the research objectives, as mentioned in section 4.3.1. By way of comparison, Krueger and Summers (1986) used fewer than 100 for small firms and more than 1,000 employees for big firms. However, at this stage, because there are no available data permitting us to freely choose the number of employees, and because such definitions are the official definitions used in Saudi Arabia by GASTAT, and because this study is the first attempt of its kind to estimate the elasticity of substitution in different sizes of Saudi industries, it can be argued that these estimations should be regarded as initial estimations only, and further studies should be conducted with a higher number of employees as a definition of big firms. Another plausible reason for the failure to reject these null hypotheses (in 7 and 8) is that the capital variable in both industries is the same because capital is reported as an aggregate measurement, whereas it should be higher (more intensive) in the big industry than in the micro industry. In other words, the dominating micro/informal industries in the economy might have downwardly affected the measurement of the (aggregated) average capital.

Although the null hypothesis in 7 and 8 were both rejected, the confidence intervals of the elasticity estimations in both industries indicate that there may complementary relationships. This is because the overall range estimates of direct elasticity in micro and big industries are (-7.1; 16) and (-0.18; 0.55), respectively, whereas the range estimates of partial elasticity in micro and big industries are (0.33; 0.77) and (0.51; 0.6), respectively. In both industries, the estimates of direct elasticity of substitution have interval estimates containing zero; hence, it cannot be ruled out that the relationship between Saudi labor and capital is a perfect complementary relationship. Also, in both industries, the estimates of partial elasticity of substitution also have interval estimates greater than zero and less than one; hence, there is evidence in favor of a complementary relationship between the composite inputs and foreign labor (i.e., they have very low degrees of substitution).

Besides, although the null hypotheses above have been evaluated according to the research procedure in economics and other social sciences, there is growing concern among statisticians that this procedure may be considered bad practice, leading to the misinterpretation of statistical tests. This is because the meaning of the p-value could be lost or misunderstood by representing it as a dichotomous value, which would make findings either significant or insignificant. For instance, Greenland et al. (2016) state that:

> the P-value is then the probability that the chosen test statistic would have been at least as large as its observed value if every model assumption were correct, including the test hypothesis. … Observing P > 0.05 for the null hypothesis only means that the null is one among the many hypotheses that have P > 0.05. Thus, unless the point estimate (observed association) equals the null value exactly, it is a mistake to conclude from P > 0.05 that a study found “no association” or “no evidence” of an effect. (pp. 339–342)

In short, a small or large null p-value means that the data may or may not be considered as unusual if all assumptions related to the statistical model and the null hypothesis are correct (Greenland et al., 2016).

5.4 Implications of the Study

This study has five implications, summarized in the following sections.

First, this study theorized and estimated the impacts of micro stores in Saudi Arabia—the first academic endeavor of its kind. Despite the dominance of micro stores in the Saudi economy since the 1970s, there appear to be no previous studies examining micro stores and their impacts on market structure, and thereby on (un)employment. Thus, this research has drawn attention to the economic phenomenon of micro stores by investigating the Saudi labor market from the perspective of industrial organizations. The study examined the Saudi economy over the past 50 years, revealing important characteristics of the local labor market. These characteristics include the structure of employment, wage

levels, and labor share. A comparison was also drawn between Saudi and non-Saudi labor and the distribution of variant establishment sizes. The study then examined the differences between micro and large establishments in terms of elasticity of substitution between Saudi and non-Saudi labor, among other aspects of working conditions.

Unlike mainstream economists and policymakers, who advocate small-business models as a solution for Saudi unemployment, this study suggests that the status quo of the micro store model is likely to distort the labor market and may hinder the performance of the private sector, thereby increasing the unemployment of the local workforce. The micro stores model is a unique economic phenomenon; therefore, it cannot be compared to the small business model usually discussed in the literature. Micro stores operate mostly as distribution channels for imported goods and provide services in various sectors, although they concentrate on a few specific sectors. Because micro stores are almost solely operated by foreign labor, these sectors risk lack of supply of foreign workers as a result of localization policies or other exogenous shocks. Moreover, because micro stores are concentrated in few industries, they have not contributed to economic diversification. In addition, the phenomenon of “commercial concealment” is usually pervasive in micro stores because many people use these stores to generate additional fixed income by renting them out to foreign workers. Therefore, the micro stores model can be considered a parasitic model. Over the past 50 years, it has not contributed to creating suitable jobs for Saudis nor has it enhanced product markets, which would have allowed the private sector to contribute to economic growth and diversification.

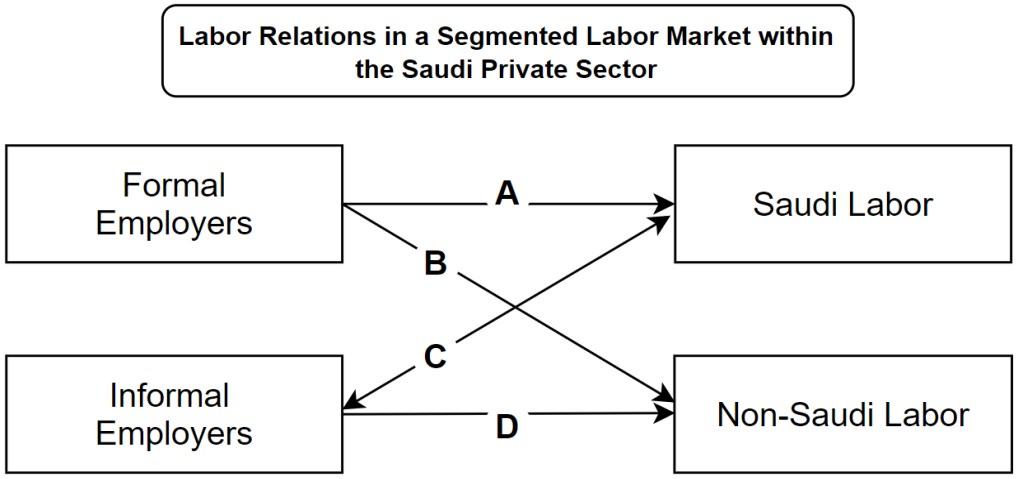
The fact that both the rate of micro store jobs and local unemployment have increased in recent decades leads us to question whether the micro stores model generates meaningful job opportunities for the growing local workforce. Recent estimates indicate that approximately 12% of the total local workforce is unemployed and that 74% and 54% of this unemployed workforce are in the 15 to 29 and 20 to 29 age groups, respectively (General Authority for Statistics, 2019a). Given that the young Saudi population constitutes a large proportion of

the total population, this may intensify the problem of unemployment and deepen despair among the Saudi youth, to the extent of threatening national stability. This concern has been raised by other authors, especially after the so-called “Arab Spring” (Leber, 2019). Unemployment can be exacerbated by several factors, for example, (i) technological progress facilitates the shift toward capital-intensive production and automation; (ii) international trade (despite its benefits) can result in operations being outsourced abroad when firms move to other countries; (iii) the number of new entrants into the labor force is still high and new labor groups (e.g., female labor and previously discouraged labor) have entered the labor market; (iv) there are reduced or insufficient new job openings in the public sector compared with the number of new labor entrants, particularly those still preferring to work in the public sector as opposed to the private sector; and (v) labor regulations aiming to increase employment in the private sector (e.g., minimum wage and localization policies) may reduce the number of existing businesses or deter new ones from entering the market. Therefore, policymakers may need to reconsider both the micro stores model as well as regulations likely to have counterproductive effects.

Second, this study has used the theory of labor market duality or segmentation in the analysis of Saudi unemployment, thereby providing new insights into the Saudi market structure. This has several theoretical implications. Previous studies have mentioned the duality between the public and private sectors. However, this research identifies a different labor market duality, namely, the duality within the private sector itself between micro and large establishments. These two different sizes of establishments represent two types of labor markets. The former represents an informal, underdeveloped labor market, while the latter represents a formal, developed labor market. These have crucial impacts on the employment and participation of Saudi labor. Therefore, this study argues that the dominance of micro stores in the private sector has created an informal labor market with poor working conditions and a less diversified economy. This could be the key reason explaining the low levels of Saudi recruitment and participation in this sector, thereby increasing the likelihood of unemployment in the sector.

In addition, using the theory of dual labor market in the analysis of the labor market assists us to evaluate different public policies, such as spending on education and training, subsidies for the private sector, and localization policies. For example, according to the dual labor market theory, the return on human capital in the secondary labor market is less than in the primary labor market (Doeringer & Piore, 1970; Osterman, 1975; Piore, 1968). As argued in this study, micro stores are the chief contributors to the issue of the dual labor market in the private sector, which has various implications. To illustrate the labor duality within the private sector, Figure 5.2 shows two different types of employers facing two different types of labor pools as they operate in a dual labor market.

Figure 5.2: Labor Relations in a Segmented Labor Market



Source: Compiled by the researcher

As can be seen in Figure 5.2, the first employer is formal, for example, large and listed companies. This employer faces two types of labor pools: Saudi and non-Saudi laborers, represented by the arrows A and B. Formal employers usually recruit Saudi labor for high- to mid-ranking jobs, although more so for mid-ranking jobs. This is because high-ranking jobs require specific skills or expertise (which Saudi labor may not have), while the low-ranking jobs are shunned by Saudi labor. This is also because the mid-ranking jobs are typically more suitable for Saudi skills. In addition, large employers recruit non-Saudi labor for all types of jobs, particularly for high- and low-ranking jobs.

As also shown in Figure 5.2, the second employer is informal, for example, micro and small businesses. This employer also faces two types of labor pools: Saudi and non-Saudi laborers, represented by the arrows C and D. However, in this case, Saudi labor can be considered self-employed because they are the owners of the micro stores, hence, the double arrow C. Because a large percentage of Saudi micro store owners does not work in the stores, these owners tend to sublease the stores to foreign labor, as discussed in section 3.2.3. Thus, labor demand and supply in these two labor markets is considerably different.

In the formal labor market, labor demand for Saudi labor can be induced by different job rankings in different industries. In the informal labor market, by contrast, there is no variety of job rankings or professions within an establishment. Because of the contractual relationship between Saudi and non-Saudi labor in micro stores, it can be argued that increasing the number of micro stores increases the number of foreign laborers. In fact, micro stores create an artificial shortage of Saudi labor, perpetuating the dependence on cheap foreign labor. Consequently, localization policies may succeed in mid-ranking jobs in the formal labor market, but not in the other cases. This is because substituting foreign labor for Saudi labor can be easier in mid-ranking jobs than in high- and low-ranking positions, whereas substituting foreign labor for Saudi labor in the informal labor market is highly unlikely. However, given the small percentage of formal employers, mid-ranking jobs are scarce in the private sector. Therefore, the level of Saudi employment is low compared with the public sector. Because foreign labor fulfills crucial tasks in high- and low-ranking jobs, enforcing localization policies may harm the production process and lead to other issues, such as shortage of supply or inflationary situations. Therefore, considering the private sector as a segmented labor market, consisting of micro stores (informal sector) and large firms (formal sector), may draw attention to this economic phenomenon. This would create different policy implications for localization policy, training, and fostering different business models.

Third, the Saudi government aims to reduce its dependence on foreign workers as well as the oil industry by developing other industries and moving toward a knowledge-based economy. This is stated in several of the country's Five-Year Development Plans as well as in the most recent economic reform, Vision 2030 (Saudi Vision 2030, 2016). Moreover, the Saudi government appears to be aware of its inability to continue as the first-and-last-resort employer for the steadily increasing local labor force. Thus, in recent years, the government has sought to increase Saudi employment in the private sector by enforcing localization policies to replace foreign workers with Saudi workers and by introducing an expatriate levy. This study argues that given the current structure of the Saudi labor market, the government is unlikely to succeed or will face major difficulties in achieving these goals. This is because the study has identified the key factors that have unintentionally distorted the structure of the Saudi labor market since the 1970s. The findings of this thesis may provide a more coherent explanation of persistent unemployment and other labor issues.

The key factors distorting the Saudi labor market stem from three policies, namely, (i) the rapid expansion of the public sector and attraction of Saudi labor, (ii) the rapid expansion of micro stores, and (iii) the sponsorship (Kafala) system. Together, these market-distorting factors have led to a structural shift, leading toward an inefficient labor market. In other words, the three policies have had unintentional consequences that have distorted the market mechanism of labor demand and supply, thereby undermining economic efficiency. Overall, while the sponsorship system has distorted the total labor supply, micro stores have distorted the product markets by adopting a low capital-intensive model and shifting the labor demand toward low-skilled foreign workers. The public sector, for its part, with its appealing employment policies compared with the private sector, has resulted in a sectoral preference against the private sector. This has, in turn, led to shocks in labor demand for Saudi labor. These were initially caused by the high demand for Saudi labor in the 1970s and early 1980s, and later, by reducing recruitment to mitigate the consequences of declining oil prices and induce employment growth of Saudi labor in the private sector.

Fourth, this study uses the Establishments Economic Survey that is annually published by GASTAT, based on the Enumerating Establishments census. This approach is adopted because, to the best of the researcher's knowledge, to date, no study has utilized this census despite its usefulness in explaining the Saudi labor market structure. This study uses the 2010 census for the descriptive analysis and the survey series (2010–2017) for the descriptive and empirical analyses. The Enumerating Establishments was the first-ever census of establishments, published in 2010 by the General Authority for Statistics (2010a).

These data reveal several important facts about the structure of the Saudi private sector and labor market, which can be summarized as follows. (i) The structure of the Saudi industry and Saudi employment distribution reveals that the least localized (i.e., least Saudized) industries are the least concentrated, while the most localized (i.e., most Saudized) industries are the most concentrated (see Table 2.8 and Table 2.9). (ii) The Saudi market structure is dominated by a high number of fragmented establishments (i.e., micro stores and small establishments constitute 97% of establishments), which is likely to hinder the transformation of the private sector into a developed, efficient sector (see Table 2.10). (iii) The median estimates clearly reveal the difference in output between micro and big industries (see Table 5.1). This is because the added value in big industries is about 16 times greater than in micro industries, even though the firm size distributions for micro and big industries are about 83% and 3%, respectively. In terms of the number of workers, big industries, by definition, have more labor (Saudi and non-Saudi). However, establishments in both micro and big industries recruit more non-Saudi workers than Saudi workers. (iv) The net capital formation is aggregated; hence, we cannot comment on the size of each industry separately. However, the median value of the net capital formation is about SR645,000 only. This indicates that the capital investment in the private sector can be considered exceedingly low.

Fifth, this study employs the nested CES production function, which enables us to estimate the direct elasticity of substitution between Saudi labor and capital, and the partial elasticity of substitution between composite inputs (Saudi labor

and capital) and foreign labor. Such an estimation provides new insights into the elasticity of substitution between inputs of production in the Saudi market. Moreover, this study estimates the nested CES production function in different labor markets to test the research hypotheses regarding labor market duality in the private sector. Researchers may continue, using the same or similar theoretical paradigm, to investigate the difference in the elasticity of substitution in different industry sizes instead of using overall estimation, which may not necessarily lead to a better understanding of the substitution relationships between Saudi and non-Saudi labor in the market.

Before discussing the implications of the estimation of the nested CES production function for micro and big industries, it is useful to recall the interpretation of different parameters. Three crucial parameters are summarized in Table 5.12 by (Heun et al., 2017). Two parameters are relevant here. First, lambda (λ) represents the constant Hicks-neutral annual rate of technological change. The lambda reflects exogenous factors influencing economic growth; in other words, when its value is large, growth is not explained well by endogenous factors of production and vice versa. Second, sigma (σ) represents the elasticity of substitution measuring who easy the substitution between two inputs. Table 5.9 and Table 5.10 show that the estimations of λ indicate that for eight years (2010–2017), micro and big industries grew annually on average at 6% and 20%, respectively. This means that the rate of technological change in big industries is about three times greater than that in micro industries. Further, it means that economic growth is more driven by big industries, rather than micro industries.

Table 5.12: Policy Implications Arising from Fitted CES Parameters

Fitted Parameter	Magnitude in Preferred Models	Interpretation	Policy Implication
$\hat{\lambda}$	Large	growth *not* explained well by endogenous factors of production	focus investment on technology and innovation to accelerate growth
	Small	growth explained well by endogenous factors of production	focus investment on endogenous factors of production to accelerate growth
$\hat{\alpha}_i$	$0 \leftarrow$	little marginal effect of increasing i, i not a significant constraint in growth	invest in factors other than i to accelerate growth
	$\rightarrow 1$	significant marginal effect of increasing i, i constrains growth	invest in i only to accelerate growth
$\hat{\sigma}_{i,j}$	$0 \leftarrow$	i, j are complements, constraints in $i \; or \; j$ will impede growth	invest to minimize exposure to low substitutability
	$\rightarrow \infty$	i, j substitutable, constraints in one of $i \; or \; j$ will not impede growth	little concern for constraints in one of $i \; or \; j$

Source: Table 4 in Heun et al. (2017, p. 25), where $\hat{\lambda}$ i is the growth rate, $\hat{\alpha}_i$ is output elasticity, and $\hat{\sigma}_{i,j}$ is elasticity of substitution.

In micro industries, the estimation of the efficiency parameter (γ) is relatively high, which may indicate efficient usage of the production inputs. In contrast, the estimation of the rate of technological change (λ) is relatively low, indicating that growth in micro industries may be explained well by endogenous factors of production. Both estimated parameters suggest that the status quo of micro stores may not require technological advancement since they meet their business purposes without the need to incur extra costs. However, micro stores may not be able to sustain any requirements (e.g., by the government) that increase their production costs, such as upgrading production technology or localization policies. In big industries, by contrast, the estimation of the efficiency parameter (γ) is higher than 1, which may still indicate efficient usage of production inputs. However, the estimation of the rate of technological change (λ) is relatively high, which indicates that growth in big industries may not be explained well by endogenous factors of production. Hence, policymakers would

need to encourage more investment in technology and innovation to accelerate growth. The difference between micro and big industries in terms of both efficiency parameters and the rate of technology changes could be attributed to the costs of Saudi recruitment and capital expenditure by the big industries. These costs exist at exceedingly low rates in micro industries. However, it should be noted that an efficient parameter is not a unitless parameter; hence, it may not be instructive to compare the efficient parameter of one industry with another. This is because it only measures the aggregate value of the output of an industry (or any other unit) to the total costs of input used to produce that output.

In terms of elasticities of substitution, Table 5.9 and Table 5.10 show that the estimations of direct elasticity of substitution in micro and big industries are 4.42 and 0.185, respectively. This means that the relationships between Saudi labor and capital tend to substitute each other in micro industries and complement each other in big industries. In contrast, the estimations of partial elasticity of substitution in micro and big industries are 0.55 and 0.556, respectively. This means that the relationships between the composite inputs (Saudi labor and capital) and foreign labor tend to complement each other in both micro and big industries. In addition, hypothesis testing (number 7 and number 8) did not support the existence of a difference between micro and big industries in terms of the estimations of direct and partial elasticities of substitution. Lever (1996) studied firm size and employment determination in Dutch manufacturing industries and examined how a firm's size may affect the determination of employment. He found that "the elasticity of factor substitution and the real wage elasticity at constant output are nearly the same for large and small firms" (Lever, 1996, p. 389). However, this study maintains that the micro stores model is a unique phenomenon that should not be compared with small firms discussed in the literature. Moreover, this study used the official measurement of big firms, which may not be comparable to other studies, as discussed earlier.

Although the null hypotheses in 7 and 8 were both rejected, the confidence intervals of the elasticity estimations in both industries indicate that there may be complementary relationships. This is because the overall range estimates of

direct elasticity in micro and big industries are (-7.1; 16) and (-0.18; 0.55), respectively, whereas the range estimates of partial elasticity in micro and big industries are (0.33; 0.77) and (0.51; 0.6), respectively. In both industries, the estimates of direct elasticity of substitution have interval estimates containing 0; hence, it cannot be ruled out that the relationship between Saudi labor and capital is a perfect complementary relationship. Further, in both industries, the estimates of partial elasticity of substitution also have interval estimates greater than 0 and less than 1; hence, there is evidence of a complementary relationship between the composite inputs and foreign labor (i.e., they have very low degrees of substitution).

In general, it can be said that all inputs of production in both micro and big industries have very low degrees of substitution. These findings corroborate prior research, cautioning against the counterproductive effects of localization policies. Several authors argue that the elasticity of substitution between Saudi and non-Saudi workers is low, making the implementation of localization policies difficult, if not impossible (Abdalla et al., 2010; Fasano & Goyal, 2004; International Monetary Fund, 1997). These findings are also in line with several studies discussed in section 2.4.2.2, cautioning against the adverse impacts of localization policies (Al-Filali & Gallarotti, 2012; International Monetary Fund, 2018; Kapiszewski, 2000; Peck, 2017; Ramady, 2010, 2013; Sadi & Henderson, 2010). For instance, Mashaal (2013) estimated the (general) elasticity of substitution between Saudi and non-Saudi labor using the CES function. He found that the elasticity of substitution was below unity at 0.457 and statically significant, which indicates the complementarity relationship between the two types of labor.

Outside the Saudi context, prior studies estimating the elasticity of substitution between labor and capital also tend to show low degrees of substitution. Chirinko (2008) examined the elasticity of substitution between labor and capital considering long-run and short-run models. He concluded that “[w]hile some estimates of sigma are above one, the weight of the evidence suggests that sigma lies in the range between 0.40 and 0.60” (Chirinko, 2008, p.

681). Pereira (2003) reviewed major studies estimating the elasticity of substitution between labor and capital from the past 40 years and found that the elasticity estimations were below 1. Hence, many previous studies found that the elasticity substitution tends to be below unity, indicating low substitution degrees or a complementary relationship between labor and capital.

Having discussed the estimations of elasticity of substitution, an important question can be asked, namely, what are the implications of low elasticities of substitution between Saudi labor and other inputs of production (i.e., capital and foreign labor) for localization policies and human capital investment? Theoretically, there is a link between own-wage elasticity and elasticity of substitution. Ehrenberg and Smith (2012) point out that own-wage elasticity is influenced by several factors. These can be summarized by the Hicks–Marshall laws of derived demand, as follows:

> These laws assert that, other things equal, the own-wage elasticity of demand for a category of labor is high under the following conditions: 1. When the price elasticity of demand for the product being produced is high. 2. When other factors of production can be easily substituted for the category of labor. 3. When the supply of other factors of production is highly elastic (that is, usage of other factors of production can be increased without substantially increasing their prices). 4. When the cost of employing the category of labor is a large share of the total costs of production. (Ehrenberg & Smith, 2012, p. 97)

Why is the own-wage elasticity of Saudi labor important? In other words, what does low or high own-wage elasticity of demand for Saudi labor imply for labor supply, labor demand, and policymakers? Low responsiveness of labor supply and demand means that existing policies, such as localization or minimum wage policies, may be ineffective. Table 5.13 summarizes the effects of two key policies aiming to increase Saudi employment in the private sector, namely, the minimum Saudi wage and the minimum Saudi employment level (i.e., Nitaqat policy).

Table 5.13: Policies Aiming to Increase Saudi Employment in the Private Sector

Key Policies	Scale and Substitution Effects	Likelihood and Constraints
Minimum wage of Saudi labor	**Scale effect**: Increasing wage may induce firms to reduce number of Saudi employees.	The reduction depends on several factors: (i) the own-wage elasticity of demand for Saudi labor, and (ii) firms being constrained by the Nitaqat requirements of Saudi employment.
	Substitution effect: It may induce firms to substitute Saudi labor for other inputs (capital and foreign labor).	Substituting Saudi labor for other inputs is constrained by several factors: (i) the low degree of elasticity of substitution, (ii) the low supply elasticity of other inputs, and (iii) the low capacity of scale and scope of economies for most establishments.
Minimum levels of Saudi employment (localization/Nitaqat policies)	**Scale effect**: It may induce firms to increase Saudi employment at least to the minimum Nitaqat required levels.	Firms face three main challenges: (i) finding enough Saudi labor willing to work for the minimum or average wage, (ii) finding enough Saudi labor suitable to take over foreign labor jobs, and (iii) paying the minimum wage or even higher to attract Saudi labor.
	Substitution effect: It may induce firms to substitute foreign labor with Saudi labor and to spend more on capital and invest in technology.	Firms may be constrained by the low supply elasticity of other inputs (capital and foreign labor), which reduces their ability to expand and employ more capital and foreign labor.

Source: Compiled by the researcher

When the Saudi wage level increases (e.g., by minimum wage legislation), this will result in scale effects and substitution effects. For the scale effect, the increase in the Saudi wage level would, ceteris paribus, induce firms to reduce the number of Saudi employees. This reduction of Saudi employment would depend on the own-wage elasticity of demand for Saudi labor. If the own-wage elasticity were elastic, this would mean that the relative change in wage would be less than the relative change in Saudi employment. This would decrease a firm's expenditure on employment, and consequently, the aggregate earning of Saudi labor would decline. If the own-wage elasticity were inelastic, this would mean that the relative change in wage would be greater than the relative change in Saudi employment. This would increase a firm's expenditure, and

consequently, the aggregate earning of Saudi labor would increase. The reduction of Saudi employment also depends on the localization requirements. In other words, firms with Saudi employment levels that are already higher than the localization requirements may be able to reduce or be discouraged from recruiting more Saudi employees. In contrast, firms with Saudi employment levels that are lower than the localization requirements would face difficulties in meeting both minimum wage level and Saudi employment requirements.

For the substitution effect, the increase in the Saudi wage level would, ceteris paribus, induce firms to substitute Saudi labor for other inputs (capital and foreign labor). This substitution would depend on several factors: (i) the elasticity of substitution, (ii) the supply elasticity of other inputs, and (iii) the capacity of scale and scope of economies for most establishments. As the empirical estimations reveal, the elasticity of substitution between Saudi labor and other inputs (capital and foreign labor) tend to be below unity, indicating complementary relationships. In other words, the utilization of inputs goes hand in hand, so their utilization either increases, reduces, or tends to go in the same direction. Hence, the reduction of Saudi employees resulting from an increase in their wage levels may also induce firms to decrease their investment in capital and foreign labor, thereby further reducing job creation. In addition, this substitution depends on the supply elasticity of other inputs, which is likely to be low (inelastic). Capital supply elasticity may be inelastic because access to credit may be possible for large firms, while micro stores and other similar stores would be unlikely to have credit access. Even if they could access credit, the capital substitution may not be applicable to micro stores given their structure and production processes. Foreign labor supply elasticity may be inelastic because access to foreign labor is limited or granted only to firms complying with localization requirements. Finally, this substitution depends on the capacity of scale and scope of economies for most establishments. Because they are dominated by micro stores, Saudi establishments do not have economies of scale and scope, enabling them to substitute between production inputs.

The discussion now turns to the second policy, as shown in Table 5.13. Enforcing the localization policy will result in scale effects and substitution effects. For the scale effect, enforcing localization policy would, ceteris paribus, induce firms to recruit more Saudi labor—at least to the minimum Nitaqat required levels. Localization policies assume that the private sector is well functioning and able to substitute foreign labor with Saudi labor. However, this assumption is questionable because firms face at least three main challenges. The first challenge is finding enough Saudi labor willing to work at the minimum or average wage levels. In other words, the supply elasticity of Saudi labor with respect to wages in the private sector tends to be inelastic. The second challenge is finding enough Saudi labor suitable to take over foreign labor jobs. In other words, the elasticity of substitution between Saudi and non-Saudi labor is low, indicating a complementary, not substitution, relationship. The third challenge is the ability of firms to pay the average or even the minimum wage level to attract Saudi labor. Consequently, compliant firms may become less competitive. Moreover, the allocation of Saudi workers may be inefficient, particularly when they are recruited to unproductive positions or as bogus localization. In contrast, non-compliant firms may downsize to a quota that requires fewer Saudi employees or relocate their businesses outside the country to avoid paying taxes for being non-compliant with localization legislation.

For the substitution effect, enforcing localization policy would, ceteris paribus, induce firms to substitute foreign labor with Saudi labor. However, because the relationships between Saudi labor and other inputs are complementary, it is expected that increasing the level of Saudi employment would induce large firms to utilize more capital (e.g., enhance technology) and foreign labor. Nonetheless, firms may be constrained by the low supply elasticity of other inputs (capital and foreign labor), which would reduce their ability to expand and employ more capital and foreign labor given the structure and distribution of establishment sizes in the private sector.

Of course, the net effects of both policies (i.e., minimum wage and Nitaqat) may be opaque. This is because a whole set of factors is happening at the same

time by which the determination of the dominated (either scale or substitute) effects is not clear. Overall, however, because most products in the private sector are produced/operated by foreign labor, the requirement of Saudization and minimum wage laws may increase the costs of production. In turn, this is likely to lead to “production bottlenecks” and inflationary situations, especially in the short run. In the long run, local producers may have to adjust their production process by integrating Saudi labor, enhancing technology, or seeking outsourcing from abroad. Otherwise, they would face the risk of bankruptcy. Consequently, the economy may face capital flight and/or decrease/deter foreign direct investment. This would reduce Saudi employment levels in the private sector and hinder economic growth. Therefore, localization and minimum wage policies may have counterproductive effects.

Against this backdrop, three potential scenarios could face the private sector in the future. The first is the continuation of micro stores as the main business model. Under this model, the private sector would continue as an informal sector that is critically dependent on foreign workers and unable to transform into a developed one. Over time, more and more firms may go out of businesses or relocate to nearby countries. The second scenario is increasing the number of foreign companies that penetrate the local market. This would bring about several benefits for the economy, including increasing Saudi employment in the private sector and increasing tax revenues. However, foreign investment is usually attracted by incentives, as a result of which, some localization and/or taxation requirements may be relaxed. Under this model, profits generated in the domestic economy would be transferred to parent foreign companies. Eventually, they would displace many national businesses. The third scenario is creating joint ventures between the government and the private sector to develop certain industries while recruiting and training local labor. This directed approach is recommended in this study (as discussed in section 3.4) because it would accelerate transition to a more developed and formal economy.

Having discussed the implications of this study on the localization policies, the discussion now turns to the implications of this study on the human capital theory.

The human capital theory is frequently used in economic studies investigating the issue of Saudi unemployment. Researchers usually investigate whether the high Saudi unemployment rate can be attributed to insufficient skills of Saudi labor. Researchers using this theory tend to conclude that Saudi labor lacks the skills required by the labor market. However, the main problem in applying human capital theory to the Saudi labor market is that proponents of the insufficient skills hypothesis implicitly assume that the Saudi private sector is well functioning. This study rejects this notion, and attempts to provide a more coherent analysis of the key market-distorting factors that hinder the transition of the economy to a more formal and efficient one. It seems that many researchers and employers have high expectations of the education system, to the extent that they expect graduates to be “instant specialized workers” without any need for further (specific) training. Thus, they tend to overlook the fact that the market structure itself has an impact on human capital investment.

The empirical findings of this study reveal that the relationship between Saudi labor and foreign labor in both micro and large industries is a kind of complementary relationship, but not a substitution one. This can be attributed to the fact that the different labor groups (i.e., Saudi and non-Saudi labor) have different characteristics, which means they are recruited for different, complementary tasks. Skill level is one of the characteristics that varies among Saudi and non-Saudi labor, although it may not be considered the most important one. The skill level is the focus of the human capital theory. Hence, researchers tend to stigmatize the education system for failing to equip the new generation with the right skills to fulfill job requirements. In contrast, this study argues that the structure of the Saudi market is the main obstacle to the development of human capital. This can be attributed to several reasons.

First, this study has found that foreign labor is relatively less skilled than Saudi labor based on years of schooling. Second, this study has also found that the unemployment rate among skilled Saudi labor more than unskilled ones such as those who graduate from science disciplines and abroad universities. Finally, as mentioned in section 2.4.1.2, there are at least three important factors that affect human capital investment: (i) diversification of the economy, (ii) size of the economy, (iii) distribution of firm size. The findings of this study reveal that the Saudi market is undiversified, small, and dominated by micro and small establishments. Consequently, these combined factors have largely contributed to the issue of low human capital investment in specific training through on-the-job training programs. For instance, this study found that the least concentrated industries (e.g., the wholesale industry) tend to be dominated by micro stores, require no specific skills, and recruit unskilled cheap foreign labor. In contrast, the most concentrated industries (e.g., mining) tend to be dominated by large firms, require specific skills, and recruit skilled Saudi labor (see sections 2.2 and 2.4.1.2). This fact alone provides enough evidence to refute the hypothesis of insufficient skills of Saudi labor. This finding can be attributed to the reality that recruiting and training Saudi labor is costly while micro and small establishments do not generally require specific training. In cases where they do require some sort of training, they do not have economies of scale and scope to reduce the average costs of recruiting and training Saudi labor.

To conclude, the human capital theory can have useful applications for the Saudi labor market. However, the crucial role played by the human capital theory cannot be overestimated. As affirmed by Borjas (2003), “the human capital literature emphasizes that schooling is not the only—and perhaps not even the most important— determinant of a worker’s skills” (p. 1339). Therefore, the focus of policymakers and researchers should be not only be on how to make the national labor force more competitive to increase their chances of recruitment (labor supply), but also on assisting the private sector to create meaningful jobs as well as recruiting and training national labor (labor demand). This would enhance voluntary transactions between employers and national labor. The argument of insufficient skills of Saudi labor as the main explanation of the

continuous high unemployment rate among Saudis may not be convincing for several reasons. First, this argument seems to implicitly assume that the Saudi private sector is functioning well, with professional firms creating meaningful jobs, and its only problem is the shortage of a well-trained labor force. This study contests such an assumption by evaluating the market structure and pointing out that the prevailing business model (i.e., micro stores) may contribute to slowing down employment growth and human capital development of Saudi labor. Second, focusing only on labor supply theories (e.g., human capital theory), while neglecting labor demand (e.g., market structure and working conditions) has been proven to be an inadequate approach. This is especially so, given the issue of educated unemployment, underemployment, and low labor participation; hence, the inefficiency of resource allocation. Thus, with the underdeveloped private sector, the greater the educational attainment, the lower the expected participation, which exacerbates educated unemployment and/or underemployment in the public sector. Third, drawing on questionnaires to test whether a skill mismatch hypothesis exists may be unreliable because of the likelihood of bias stemming from measurement errors, as discussed in section 2.4.1. Finally, although the argument of the skill mismatch hypothesis has partial validity, it has left some questions unanswered. These include the following:

- If the problem is the insufficient skills of Saudi labor, how can we explain the positive correlation between educational attainment and unemployment rates over the past decades? Moreover, how we can explain the dominance of low- and semi-skilled foreign labor in the private sector while Saudi labor has higher educational attainment?
- If the problem is the low quality of the education system, how we can explain unemployment among job seekers who trained and graduated from overseas universities?
- Given the fact that both education and on-the-job training are two forms of human capital investment, why do the proponents of human capital theory emphasize education while overlooking on-the-job training despite its important role in transferring expertise? Moreover, if the education system failed to train the labor force, why we do not see the private sector offset such shortcomings by on-the-job training, as the private sector has done in other countries?

- The Saudi government strives to improve Saudi human capital through various programs, such as enhancing the quality of education, providing scholarships to pursue higher degrees or obtain proper training abroad, training unemployed people as well as offering various forms of subsidies to employers recruiting Saudi workers, such as bearing the whole or part of Saudi wages over a certain period. Yet the general response from the private sector has been disappointing. Why is this so?

In Chapter 3, the study attempted to provide an alternative explanation for Saudi unemployment and other low productivity symptoms of the Saudi private sector, which may give some answers to the above questions. To recall, Chapter 3 discusses three market-distorting factors: (i) the rapid expansion of the public sector and attraction of Saudi labor, (ii) the rapid expansion of micro stores and the disguised shortages of the national workforce, and (iii) the sponsorship system. Those factors have arguably undermined the efficiency of the Saudi labor market, leading to a segmented labor market and ultimately, persistent unemployment. Together, these market-distorting factors have created a structural shift, leading toward an inefficient labor market. In other words, the three policies have had unintentional consequences that have distorted the market mechanism of labor demand and supply, thereby undermining economic efficiency. Overall, the public sector expansion and attraction of Saudi labor since the 1970s has resulted in an increase in Saudi employment. Because of the higher demand in the public sector for Islamic and managerial studies graduates compared with other specializations, there is a high proportion of Saudis who pursue such programs. However, when the public sector became saturated, educated unemployment increased among all specializations. Meanwhile, micro stores distorted the product markets by adopting an exceedingly low capital-intensive model and shifting the labor demand toward low-skilled foreign workers. Because of the dominance of micro stores, the private sector became largely peripheral, lagging behind in development, lacking diversity, and incapable of recruiting and training relatively skilled Saudi labor. The sponsorship system, for its part, distorted the total labor supply, perpetuating dependence on cheap foreign labor.

Therefore, market-distorting factors are important in the analysis of the Saudi labor market because they help to explain at least seven central issues: (i) the high and persistent unemployment rate, (ii) the low labor participation rate among Saudi men and women or the slow process of integrating local labor into the private sector, (iii) the low human capital investment in specific skills through on-job-training programs such as internships and apprenticeships, (iv) the poor economic performance of the private sector, why it is susceptible to localization policies and any changes in government programs (subsidies and purchases), (v) underemployment and low labor productivity in the public sector and its low resilience in the face of economic shocks resulting from fluctuations in oil prices, (vi) high dependence on foreign labor in all industries, and (vii) the slow process of economic transformation and diversification.

5.5 Conclusion

The main ideas discussed in this chapter can be summarized as follows:

- The descriptive statistics reveal three key features (see Table 5.1). First, the difference in output between the two industries can be seen by looking at the median estimates. This is because the added value in big industries is about 16 times more than that of micro industries, despite the fact that the firm size distributions for micro and big industries are about 83% and 3%, respectively. Second, in terms of the number of workers, big industries, by definition, have more labor (Saudi and non-Saudi). However, establishments in both micro and big industries recruit non-Saudi more than Saudi labor. Finally, the net capital formation is aggregated; hence, we cannot comment on each industry's size separately. However, the median value of the net capital formation is only about SR645,000, which reveals how low the capital investment is in the private sector.
- The empirical findings can be summarized as follows. First, the range estimates of direct elasticity in micro and big industries are (-7.1; 16) and (-0.18; 0.55), respectively, whereas the range estimates of partial elasticity in micro and big industries are (0.33; 0.77) and (0.51; 0.6), respectively. In both industries, the estimates of direct elasticity of substitution have interval estimates containing 0; hence, we can infer that the relationship between Saudi labor and capital is a perfect

complementary relationship. In both industries, the estimates of partial elasticity of substitution also have interval estimates between 0 and 1; hence, we can infer that the relationship between Saudi labor and capital on the one hand, and foreign labor on the other, is complementary (i.e., there are low degrees of substitution). Second, the findings suggest that there may be no difference between micro and big industries in terms of substitution degree between the two groups, which is contradictory to research hypotheses 7 and 8. One likely explanation is that GASTAT's official classification of big firm size was not the best representative group for big firms. However, at this stage, because there are no available data permitting us to freely choose the number of employees, and because such definitions are the official definitions used in Saudi Arabia by GASTAT, and because this study is the first attempt of its kind to estimate the elasticity of substitution in different sizes of Saudi industries, it can be argued that these estimations should be regarded as initial estimations only, and further studies should be conducted with a higher number of employees as a definition of big firms. Another plausible reason for the failure to reject these null hypotheses (in 7 and 8) is that the capital variable in both industries is the same because capital is reported as an aggregate measurement, whereas it should be higher (more intensive) in the big industry as opposed to the micro industry. In other words, the dominating micro/informal industries in the economy might have downwardly affected the measurement of the (aggregated) average capital.

- Finally, these estimations of the elasticity of substitution in different-sized industries could still be useful for empirical studies as well as for policy. In terms of empirical studies, researchers could use the same or similar theoretical paradigm to investigate the difference of the elasticity of substitution in different industry sizes instead of relying on the general elasticity of substitution using aggregate data, which may not necessarily lead to a nuanced understanding of the substitution relationship between Saudi and non-Saudi labor. In terms of policy implications, the low degree of substitution in both industry sizes implies that the localization policies attempting to replace foreign labor with Saudi labor are ineffective.
- There are at least three important factors that can affect human capital investment: (i) diversification of the economy, (ii) the size of the economy, and (iii) distribution of firm size. The findings of this study have revealed that the Saudi market is characterized as undiversified, small, and dominated by micro and small establishments. Consequently, these combined factors have contributed largely to

the issue of low human capital investment in specific training through on-the-job training programs.

6. Conclusion

The Saudi economy shows poor performance in various sectors. The public sector is not only considered a saturated sector and suffering from underemployment and low productivity but also its revenues are largely dependent on depleted and price-fluctuating natural resources. Therefore, the sector is susceptible to recurring economic shocks. In contrast, the private sector is impaired because of its critical dependence on government support and cheap foreign labor. Moreover, limited capital-intensive investment negatively affects technological diffusion, leading to poor working conditions. Consequently, the participation and employment rates of national labor are among the lowest in the world, particularly for women. Overall, the economy suffers from high structural unemployment rates, a lack of economic diversification, and low resilience to fluctuations in the price of natural resources. Therefore, it can be argued that the private sector has never played a significant role in recruiting Saudi labor, let alone playing a leading role in diversifying the economy.

Structural unemployment refers to the skill mismatch hypothesis as one of the explanations of unemployment, when the gap between “skills acquired” (supplied) and “skills required” (demanded) widens in the labor market because of technological changes or structural shifts. This usually occurs as a consequence of technological advances, which prompt changes in the industrial structure, resulting in a shift in the skills required. In some sense, certain skills become obsolete or in lesser demand; hence, part of the labor force becomes unemployed. However, Saudi Arabia has been experiencing an interesting phenomenon since the1970s. That is, in general, the Saudi labor force has become relatively skilled while the private sector has failed to keep up with technological advances. Consequently, the public sector has become virtually the sole recruiter of Saudi labor while the private sector is incapable of generating meaningful jobs for the increasing labor force. Hence the Saudi labor market is, as this study argues, distorted by a structure impeding its functionality. This

distortion is so severe that the implications of some labor theories (e.g., the human capital theory) and public programs to correct the labor market have become inapplicable or ineffective.

6.1 Research Objectives

The primary objective of the study was to examine the employability of Saudi labor and investigate the condition of the Saudi labor market. This main objective was approached by attempting to answer the following questions:

- Why is unemployment high and persistent among Saudi labor?
- What is the role of the public and private sectors in employment growth?
- What are the market-distorting factors that could be the fundamental reasons for unemployment persistence over the last several decades despite the government's generous support to education and businesses to increase Saudi employment in the private sector?
- How has the micro stores model, as one of the market-distorting factors, contributed to the so-called dual labor market? And how does it differ from the small and medium enterprises?
- What should be done to reverse the effects of the market-distorting factors and correct market failure, thereby creating meaningful job opportunities for the increasing local labor force?
- Does the elasticity of substitution between Saudi and foreign workers in micro industries differ from that in big industries?
- What are the possible implications of the elasticity of substitution on localization policies and Saudi employment?

6.2 Summary of Prior Studies

Given the poor performance of the Saudi private sector, which is highly visible, both policymakers and researchers have endeavored to find ways to mitigate the persistent unemployment and low participation and employment rates of the local labor force. In terms of policy, the government has enforced localization (Saudization) policies, introduced the expatriate levy, and launched various subsidy programs—among other desperate measures—to integrate Saudi labor into the private sector and reduce unemployment. However, the response from private employers and national labor can be considered as falling below expectations. In terms of prior research, most studies have focused on three main theories or aspects, namely, human capital theory, institutional theory, and socioeconomic theory, or factors that affect the labor decisions of individuals (employers and employees).

The first aspect discussed in the literature is human capital theory. This theory has been frequently adopted in numerous studies investigating Saudi unemployment. One issue discussed pertaining to Saudi human capital is whether Saudi labor has “sufficient skills” as required by the labor market. Several authors concur that Saudi workers do not have the skills required by the private sector (Al-Asfour & Khan, 2014; Baqadir et al., 2011; Harry, 2007; International Monetary Fund, 1997). The common perception of private employers is that technical education fails to offer Saudi students sufficient vocational training to the level that employers require or expect. Another issue discussed pertaining to Saudi human capital is the quality of the education system and its relevance to labor market requirements. Several researchers call these factors into question, calling for the improvement of syllabi, the provision of labs, and fostering innovative thinking and problem-solving skills rather than teaching subjects that are unrelated to labor market requirements, such as Islamic and humanity studies (Mishrif & Alabduljabbar, 2018; Niblock & Malik, 2007; Sadi & Henderson, 2010). The researcher shares the same concerns, especially regarding the improvement of school buildings and students’ skills as well as integrating the English language, computer skills and programming into

syllabi. Nonetheless, there have been attempts to downplay the importance of Islamic studies, which could be criticized for various reasons, as discussed earlier. A final issue discussed pertaining to Saudi human capital theory is whether the Saudi human capital investment emphasizes higher education at the expense of vocational education. In contrast to proponents of higher education and scholarship programs to study abroad as the solution of Saudi unemployment, some authors argue that such human capital investment may contribute to increasing educated unemployment and underemployment in the public sector (Ramady, 2014; Stevens, 1986).

The crucial role played by human capital theory cannot be overestimated. The argument of the skill mismatch hypothesis as the main explanation of the persistent high unemployment rates among Saudis is not a convincing argument for several reasons. First, the skill mismatch hypothesis seems to implicitly assume that the Saudi private sector is well-functioning and run by professional firms creating meaningful jobs while its only problem is the shortage of a well-trained labor force. This assumption is questionable. The present study contests this assumption by evaluating the market structure and pointing to the predominant business model behind the sluggish employment growth and human capital development. Second, emphasizing education (supply side) without improving the private sector and increasing job creation (demand side) has led to more educated unemployment and resource misallocation. This is because, given the underdeveloped private sector, the greater the educational attainment of the Saudi labor force, the lower the expected participation, thereby exacerbating educated unemployment and/or underemployment. Third, relying on questionnaires to test whether the skill mismatch hypothesis exists could be unreliable because of the likelihood of bias stemming from measurement errors. Finally, although the skill mismatch hypothesis has some validity, it leaves certain questions unanswered, thereby casting doubt on this hypothesis as the main explanation of unemployment persistence.

The second aspect discussed in the literature is the institutional theory. Chapter 2 contained a review of three aspects of labor institutions related to the

Saudi labor market, namely, Saudi labor law, localization policies, and labor market segmentation. Prior studies have investigated different areas of Saudi labor law. One such area is whether there is legal discrimination between public and private sector workers or between Saudi and non-Saudi labor. It can be argued that apart from localization policies, there is no explicit discrimination in Saudi labor law between workers in the two sectors or between Saudi and non-Saudi labor. However, two main features could be considered as discrimination between Saudi and non-Saudi labor, namely, wages and employment eligibility. The former refers to minimum wage regulations and wage structure, which could be relevant to the public sector more than the private sector because the private sector has no unifying wage structure. The latter stems from localization policies. Moreover, there are several advantageous features of the public sector resulting from the structure of public institutions, rather than from Saudi labor law itself.

In terms of localization policy, the broad argument is that although enforcing localization policies could succeed in integrating national labor into the private sector in the short run, it could result in negative consequences in the long run. These negative consequences include underutilization of Saudi labor by placing Saudi workers in meaningless jobs, the bankruptcy of some businesses, and discouragement of the competitiveness of the Saudi market at both domestic and international levels.

In terms of the segmentation of the Saudi labor market, this factor can be viewed from several perspectives, such as the public versus the private sector, nationals versus non-nationals, skilled labor versus unskilled labor, and male versus female. There are several factors that may have contributed to the Saudi labor market segmentation: (i) the inter-regional migration of Saudis; (ii) the adoption of a temporary immigration policy (the guest worker model); (iii) the superiority of the public sector because of its ability to provide good working conditions; and (iv) the mismatch of skills between Saudi labor and the requirements of the private sector, particularly in both upper-end and lower-end jobs. In general, the Saudi labor markets face imbalanced demand and supply of labor because of labor immobility between markets. Three factors contributing

to labor immobility are occupational, geographical, and industry-specific skills. While these factors create labor shortages in some markets, they create surplus labor in others. The existence of labor immobility could be contributing to labor market segmentation, thereby perpetuating unemployment.

The final aspect discussed in the literature is socioeconomic theory or factors that may affect the labor decisions of individuals. Eight such factors are discussed in Chapter 2. (i) Wealth or non-labor income: Some authors suggest that the high per capita income of the Saudi population and the dependence of young Saudis on their parents may play a role in their reluctance to accept entry-level jobs (Birks & Sinclair, 1979; Spiess, 2010; Wilson et al., 2012). Relatively high wealth or non-labor income of individuals on the one hand, and the low wage levels in the private sector on the other, make unemployed Saudis prefer unemployment to working at wage rate perceived to be below subsistence level. (ii) Low incentives in the private sector relative to the public sector. Because of sectoral differences, unemployed Saudis may prefer to wait for opening jobs in the higher-wage sector (Hertog, 2018; International Monetary Fund, 1997; Kapiszewski, 2000; Ramady, 2010), which is likely to increase frictional unemployment rates. (iii) The role of capital-intensive projects in employment growth is limited Looney (1988). (iv) The costs of national labor are higher than those of their counterpart foreign workers (International Monetary Fund, 1997). (v) The absence of effective internship and apprenticeship programs reflects the low experience and practical skills of Saudi workers (Baqadir et al., 2011; Spiess, 2010), which reduces their chances of finding jobs in the private sector. (vi) The rural–urban migration of Saudis or the imbalanced distribution of the population creates geographical immobility of the labor force (Al-Filali & Gallarotti, 2012; Stevens, 1986; Wilson et al., 2012). (vii) There is a negative attitude toward working in the private sector, largely because of social or religious reasons (Mellahi, 2007). (viii) The lack of work ethics among Saudi workers reinforces the reluctance of private employers to recruit them (Baqadir et al., 2011).

Previous studies on the Saudi labor market and unemployment have relative importance and potential implications for the Saudi economy. However, they

suffer from one or more of the following limitations. First, they lack consistent economic analysis explaining why unemployment persists despite government generous support for education and businesses to increase Saudi employment in the private sector and despite the improved skill attainment among Saudi labor in recent decades. Second, they focus predominantly on labor supply theories (e.g., skill mismatch hypothesis) while neglecting labor demand (e.g., market structure and working conditions). Such a narrow approach has been proven inadequate given the continued high unemployment in Saudi Arabia. Third, prior studies have not estimated the CES between Saudi labor, capital, and foreign labor in different establishment sizes. Finally, the studies have ignored the effect of micro store on (un)employment despite the prevalence of this model in the Saudi labor market. In addition, these studies have not addressed unemployment among the Saudi labor force from the industrial organization perspective despite the dramatic changes to the market structure over the past decades. It would be unrealistic to accurately address the current issues of the Saudi labor market without investigating how the labor market has evolved over the past decades.

6.3 Summary of Market-Distorting Factors

The study argues that the approaches adopted in prior literature seem to treat the symptoms instead of treating the root causes of the inefficiency of the Saudi economy. Therefore, this study proposes that three factors have had unintended consequences on the Saudi labor market. These factors have distorted the market by creating labor market duality between the public and private sector, and within the private sector, by creating duality between the formal and informal markets. These market-distorting factors are: (i) the rapid expansion of the public sector and the attraction of Saudi labor; (ii) the rapid expansion of micro stores; and (iii) the sponsorship system. Together, these factors have resulted in far-reaching economic and social effects on Saudi society. These factors can be traced back to the 1970s insofar as they are related to policies implemented during the oil boom prevalent at the time, resulting in unintended consequences

on the labor market. In other words, the three policies have had unintentional consequences that may have distorted the market mechanism of labor demand and supply, thereby undermining economic efficiency.

The first factor that has contributed to the structural shift in the Saudi labor market is the rapid expansion in the public sector and the attraction of Saudi labor since the 1970s. Over the period between 1970 and 2014, the Saudi economy has been oscillating between expansionary and contractionary fiscal policies, reflecting the fluctuations in oil price. Generally, expansionary fiscal policy was adopted over the periods of 1970–1984 and 2005–2014 while contractionary fiscal policy was adopted over the period between 1985–2004. Over the past 50 years, the state arguably has largely failed to diversify the economy to hedge against price fluctuations of natural resources. When resources increase, the state adopts an expansionary fiscal policy by expanding its provision of public goods and services, thereby increasing Saudi recruitment and undertaking new megaprojects. In contrast, when revenues decline, the state adopts a contractionary fiscal policy by ceasing the expansion of public goods and services, limiting job creation, restraining wage levels, and halting new projects. Adopting procyclical rather than countercyclical policies exacerbates economic shocks because the government sector is the main engine of economic growth through its purchases and subsidies. Hence, any reduction in government expenditure has severe repercussions on private investment and consumption.

The following factors indicate how the rapid expansion of the public sector and the attraction of Saudi labor have contributed to the creation of labor immobility, thereby leading to labor market duality and unemployment persistence. (i) The increase in the size of government and its role in the development process, particularly in the first three Five-Year Development Plans (1970–1984), has contributed to making the private sector mostly a shadow or informal sector. In contrast, Saudi employment policy has been used for decades as a response to the public sector expansions and/or used as a necessary

measure to mitigate unemployment. (ii) In the early stages of development, projects tend to be capital-intensive and long-term, which increases risks and uncertainty. At the same time, while the private sector lacks incentives to take initiatives, public expansion adopts a “project approach” rather than a “sectoral approach” (Niblock & Malik, 2007), which may explain why the private sector suffers from underinvestment, which has resulted in many economic activities being undeveloped or underdeveloped. (iii) While the public sector is able to offer decent jobs to graduates with higher education qualifications, there is a lack of professional firms in the private sector that can hire graduates with a vocational education. As a result, a negative stereotype has emerged around vocational education as leading to unemployment or low-paid jobs. Hence, the higher the level of education, the lower the likelihood of participating in the private sector because of the lack of suitable jobs (Kapiszewski, 2000; Ramady, 2010; Stevens, 1986). (iv) In the early stages of economic development, only a few cities were developed and urbanized, resulting in uneven regional development. Although it was intended to continue development in other cities, the development process stopped or stalled in the early 1980s because of declining oil revenues. As a result, there was rural–urban migration, creating an imbalanced population distribution across cities and towns (Al-Filali & Gallarotti, 2012; Stevens, 1986; Wilson et al., 2012). (v) There were crowding-out effects of government policies. One such effect stemmed from the policy of attracting Saudi labor to the private sector, especially during the first three Five-Year Development Plans from 1970 to 1984 (Al-Asfour & Khan, 2014; Birks & Sinclair, 1979; Stevens, 1986), which left the private sector without an adequate supply of national labor, and then dependent on foreign workers. Another crowding-out effect stemmed from borrowing from domestic commercial banks from about 1985 to 2003 (Wilson et al., 2012), which reduced the ability of domestic financial institutions to lend money to private investors. In the meantime, ceasing lending programs by government-owned development funds also exacerbated the financial difficulties facing the private sector at the time (Al Hajjar & Presley, 1996). (vi) While the private sector lagged in the improvement of working conditions, the public sector adopted more desirable employment policies. Therefore, the rapid expansion in the public sector and the attraction of Saudi labor contributed to sectoral

preferences, perpetuating labor market duality between the public and private sectors, and ultimately contributing to unemployment persistence among the Saudi labor force.

The second factor that has contributed to the structural shift in the Saudi labor market is the rapid expansion of micro stores. Since the 1970s, the rapid increase of city expansion, in conjunction with the urban planning system allowing the proliferation of micro stores, resulted in disguised shortages in the Saudi labor force and the emergence of informal/secondary labor markets. The following factors indicate how this occurred. (i) Increasing micro stores necessitates more expatriate workers, leading to a vicious cycle process. That is, as the number of micro stores increases, foreign labor also increases, which partially contributes to increasing micro stores even further to meet the various needs of foreign workers. This phenomenon is exacerbated by urban planning regulations, which allow micro stores to be built along commercial streets. (ii) The micro stores model can be considered a path-dependent model. This is because as long as the micro stores model is adopted, policies aimed at improving industrial standardization, implementing Saudization, or preventing commercial concealment will be ineffective. (iii) Micro stores create a disguised shortage of national labor because Saudi owners prefer not to work in those stores. Instead, they tend to use them as a passive income tool by subleasing stores to foreign workers (i.e., commercial concealment). (iv) The micro stores model, which is almost solely dependent on foreign workers, contributes to depressing wages, reducing returns on human capital, delaying the integration of the national labor force into the private sector, fostering a consumption-oriented society, and perpetuating the problem of the low productivity of the private sector (Looney, 1988; Ramady, 2013; Sirageldin et al., 1984). (v) The micro stores model can be considered an inefficient model, slowing down transformation into a more developed economy. The micro stores model could also be considered inefficient because of its limited capacity to utilize economies of scale and scope.

Overall, the good working conditions in the public sector and the poor working conditions in the private sector create labor market duality between the public and private sectors. Whereas the rationed/limited jobs of big professional firms and the abundant but degraded jobs of micro stores create labor market duality within the private sector between the primary and secondary labor markets. Moreover, while the labor shortage in the public sector was genuine, temporary, and overcome by attracting Saudi labor, the labor shortage in the private sector is continuous and mostly inflated artificially, by increasing the number of micro stores. Finally, while the public sector has expanded by essential infrastructure and Saudi employment, the private sector has expanded mostly by redundant micro stores, dependent almost solely on foreign labor. Hence, erroneously treating the micro store model prevailing in Saudi Arabia as a small enterprise model may lead to serious consequences in terms of policy implications and labor market structure. The above characteristics of micro stores result in occupational and geographical factors, reducing or even preventing labor mobility by discouraging the Saudi labor force from participating or moving within the private sector between the formal and informal labor markets. Therefore, it is argued that the rapid expansion of micro stores has contributed to the emergence of the secondary labor market, perpetuating labor market duality between the formal and informal labor markets within the private sector, and ultimately contributing to unemployment persistence in the Saudi labor force.

The third factor that has contributed to the structural shift in the Saudi labor market is the sponsorship system. This system was supposed to be a short-term, flexible policy in response to national labor constraints in the 1970s. However, it did not take long for the sponsorship system to become purely symbolic, with the sponsor’s role mostly confined to that of middleman, enabling foreign workers to enter Saudi Arabia. The sponsorship system has several loopholes distorting the labor market mechanism as well as undermining its reputation. The following factors may explain how the sponsorship system reinforced labor market duality between the primary and secondary labor markets as well as other distortive effects on the Saudi labor market: (i) it can lead to the problems of adverse selection and moral hazard because of asymmetric information among labor

relations; (ii) it undermines the free market mechanism by increasing labor frictions and inflexibility (Mahdi, 2000), which exacerbates labor immobility; (iii) it impedes economic growth by discouraging foreign and domestic investors from investing in the Saudi economy (International Monetary Fund, 2018; Kapiszewski, 2000); (iv) combined with the excessive number of micro stores, the sponsorship system reinforces the over-dependence on low-skilled foreign workers; and (v) combined with the excessive number of micro stores, the sponsorship system contributes to commercial concealment.

6.4 Summary of Methodology and Research Findings

To estimate the elasticity of substitution in different labor markets, the study focuses on two sets of establishment size, micro and big, presumed to capture the positive and negative characteristics defining these two markets. The basic premise is that these characteristics critically influence the elasticity of substitution between Saudi and non-Saudi workers. The substitutability relationship between Saudi and non-Saudi labor in different labor markets is crucial because they have critical implications in the analysis. In the primary labor market, the relationship between Saudi and non-Saudi labor can be substitutional or complementary, according to the type of skill or occupation. For example, if Saudi and non-Saudi laborers have similar skills (such as middle-level jobs), the relationship would be substitutional. Hence, replacing foreign workers would be possible because Saudi workers are willing to work in such jobs. However, if foreign workers have specific skills that Saudi workers do not possess (such as upper-end jobs) or if they hold certain jobs that Saudi workers would not accept (such as lower-end jobs), the relationship would be complementary. In the secondary labor market, by contrast, the relationship between Saudi and non-Saudi laborers is mostly complementary because of the undesirable characteristics in this type of labor market. It is also considered a complementary relationship because the micro stores model resembles the passive income model rather than an entrepreneurial one, where Saudi workers

play the role of employers (sponsors) and foreign workers play the role of employees (renters).

The study relies on the Establishments Economic Survey, annually published by GASTAT, and based on the Enumerating Establishments. The Enumerating Establishments was the first-ever census of establishments published in 2010 by GASTAT. The study examines only two establishment sizes (less than 5 and 20 or more employees) for micro and big industries to distinguish between the secondary and primary labor markets. The study also uses the nested CES production function to estimate the direct elasticity between Saudi labor and capital, and partial elasticity between composite inputs (Saudi labor and capital) and foreign labor in both micro and big establishments. The research findings are drawn from descriptive and empirical analyses.

The descriptive analysis reveals several findings, which can be summarized as follows.

1) Saudi Arabia can still be considered a rentier state that is largely dependent on revenues from natural resources. The public sector, as virtually the only source of Saudi employment, is a relatively large sector suffering from underemployment and low productivity. The private sector can be generally stratified into two parallel economies. One is formal, developed, and driven by big professional firms whereas the other is informal, underdeveloped, and driven by micro stores. The former is limited while the latter is dominant, resulting in a scarcity of meaningful jobs, lack of technological diffusion, poor working conditions, and over-dependence on guest workers.

2) The Saudi market structure is dominated by a high number of fragmented establishments (micro stores and small establishments constitute 97% of establishments), which is likely to impede the transformation of the private sector into a developed and efficient sector.

3) The structure of the Saudi industry and Saudi employment reveal that the least localized (i.e., least Saudized) industries are the least concentrated ones and vice versa, the most localized (i.e., most Saudized) industries are the most concentrated ones.

4) The Saudi labor force can be considered as semi-skilled. While middle-level jobs are scarce, Saudi workers are caught in the middle, unable to take on high-level jobs requiring specific skills nor willing to take on low-level jobs, which are abundant and mostly occupied by foreign workers.

5) The market size is relatively small and lacks incentives to invest and create meaningful jobs because of undefined property rights and high uncertainty pertaining to recurring changes in labor-related regulations and foreign labor supply.

6) The difference in output between the micro and big industries can be clearly seen by looking at the median estimates (see Table 5.1). This is because the added value in big industries is about 16 times more than that of micro industries, despite the fact that the firm size distributions for micro and big industries are about 83% and 3%, respectively. In terms of the number of workers, big industries, by definition, have more labor (Saudi and non-Saudi). However, establishments in both micro and big industries recruit non-Saudi more than Saudi labor. Finally, the net capital formation is aggregated; hence, we cannot comment on each industry size separately. However, the median value of the net capital formation is about SR645,000 only, which reveals how low the capital investment is in the private sector.

The empirical analysis reveals several findings, which can be summarized as follows.

1) The range estimates of direct elasticity in micro and big industries are (-7.1; 16) and (-0.18; 0.55), respectively, whereas the range estimates of partial elasticity in micro and big industries are (0.33; 0.77) and (0.51; 0.6), respectively. In both industries, the estimates of direct elasticity of

substitution have interval estimates containing zero; hence, it cannot be ruled out that the relationship between Saudi labor and capital is a perfect complementary relationship. In both industries, the estimates of partial elasticity of substitution also have interval estimates between zero and one; hence, there is evidence in favor of a complementary relationship between Saudi labor and capital on the one hand, and foreign labor on the other (i.e., it has a low degree of substitution).

2) The findings suggest that there may be no difference between micro and big industries in terms of substitution degree between the two groups, which is contradictory to the research hypotheses (7 and 8). One likely explanation is that the official classification of big firm size by GASTAT may not be accurately representative of big firms. However, at this stage, because there are no available data permitting us to freely choose the number of employees, and because such definitions are the official definitions used in Saudi Arabia by GASTAT, and because this study is the first attempt of its kind to estimate the elasticity of substitution in different sizes of Saudi industries, it can be argued that these estimations should be regarded as initial estimations only, and further studies should be conducted with a higher number of employees as a definition of big firms. Another plausible reason for the failure to reject these null hypotheses (in 7 and 8) is that the capital variable in both industries is the same because capital is reported as an aggregate measurement, whereas it should be higher (more intensive) in the big industry than in the micro industry. In other words, the dominating micro/informal industries in the economy might have downwardly affected the measurement of the (aggregated) average capital.

3) The estimations of the elasticity of substitution in different sizes of industry could serve as a platform for future empirical studies as well as informing policy development. In terms of empirical studies, researchers could use the same or similar theoretical paradigm to investigate the difference of the elasticity of substitution in different industry sizes instead of relying on the general elasticity of substitution using aggregate data, which may not

necessarily lead to a nuanced understanding of the substitution relationship between Saudi and non-Saudi labor. In terms of policy implications, the low degree of substitution for both industry sizes implies that the localization policies attempting to replace foreign workers with Saudi workers are ineffective.

In sum, although the public sector is still the first-and-last-resort employer of Saudi labor, its effectiveness to absorb the increasing Saudi labor force has declined as the sector has become saturated over recent decades. In contrast, the private sector has two labor markets. The primary labor market has a limited role in generating jobs because of its small size, among other reasons, while the prevailing secondary labor market is incapable of generating jobs that Saudi job seekers would be willing to accept. The inability to find jobs in the public sector or in the primary labor market increases involuntary unemployment, while unwillingness to work in the secondary labor market leads to frictional unemployment. Hence, the Saudi economy has been experiencing high unemployment rates for several decades. Evidently, the last four decades have shown the failure of the localization policies on the one hand, and the inability of the private sector to generate enough jobs on the other. This scenario requires the government to take the lead by acting as a catalyst in the country's economic development.

Two corrective measures are suggested in the thesis. First, it is suggested that the sponsorship system be replaced by the CRA. Second, it is suggested that micro stores be re-zoned or restricted within self-contained centers and that consortiums between the government and private investors be created to establish TDCs for developing certain industries and product markets while recruiting and training national labor. However, these approaches may face some difficulties. In the case of the CRA, there are many private recruitment agencies currently operating in the economy that will lose their livelihoods if the government adopts the CRA model. Hence, adopting such a model necessitates a proper mechanism to compensate their losses. Additionally, other difficulties

may face the transformative firms, including how to distinguish between different industries in which the transformative firms would operate, what industries the government should start with, and how to devise a regulatory mechanism to hold transformative firms accountable for delivering on developmental goals.

6.5 Research Contributions

This study has both theoretical and practical implications. The theoretical implications can be summarized as follows.

1) Theorizing and estimating the impacts of micro stores is the first academic endeavor of its kind, despite the dominance of the micro stores model in the Saudi economy since the 1970s.

2) Including the dual labor market theory in the analysis of Saudi unemployment provide a more nuanced understanding of the Saudi market structure, giving rise to several theoretical implications.

3) Identifying the main factors that have unintentionally distorted the structure of the Saudi labor market since the 1970s provides a more coherent explanation of persistent, high unemployment rates. Using the same or similar theoretical paradigm, researchers will be able to further investigate the difference in the elasticity of substitution in different industry sizes instead of using overall estimation, which does not necessarily lead to a nuanced understanding of the substitution relationships between Saudi and non-Saudi labor.

4) In contrast to mainstream economists and policymakers who advocate the small-business model as a solution for the Saudi unemployment, this study suggests that the status quo of the micro store model is a distorting model of the labor markets and may hinder the very performance of the

private sector, thereby increasing the unemployment of the local workforce in the KSA.

5) This study uses the Establishments Economic Survey that is annually published by GASTAT and based on the Enumerating Establishments. This approach is adopted because, to the best of the researcher's knowledge, until date no study has used the Enumerating Establishments census despite its paramount importance in enhancing the understanding of the Saudi labor market structure. This study uses the 2010 census for the descriptive analysis and the survey series (2010–2017) for the empirical analysis. The Enumerating Establishments was the first-ever census of establishments, published in 2010 by General Authority for Statistics (2010a).

6) Using the nested CES production function enables us to estimate the direct elasticity of substitution between Saudi labor and capital, and the partial elasticity of substitution between composite inputs (Saudi labor and capital) and foreign labor as well as understanding the different relationships between those inputs in a dual labor market setting.

In terms of practical implications, several reasons prompted the investigation of the impact of micro stores, which can be summarized as follows. First, the fact that both the rate of micro store jobs and local unemployment has increased in recent decades should lead us to question whether the micro stores model generates meaningful job opportunities for the increasing local workforce. Investigating micro stores draws attention to this economic phenomenon, thereby providing fresh policy implications on localization policy, training, and alternative business models.

Second, recent estimates indicate that the unemployment rate among the total local workforce is approximately 12% and that the 15–29 and 20–29 age groups constitute 74% and 54% of these unemployed locals, respectively (General Authority for Statistics, 2019a). Given the fact that the young Saudi population comprises a large portion of the total population, this may intensify

the problem of unemployment and deepen despair among the Saudi youth, to the extent of threatening the stability of the country. This concern has been raised by other authors, especially after the so-called “Arab Spring” (Leber, 2019).

Third, the Saudi government aims to reduce its dependence on foreign workers as well as the oil industry by developing other industries and moving toward a knowledge-based economy. This has been articulated in several Five-Year Development Plans and in the most recent economic reform, Vision 2030 (Saudi Vision 2030, 2016). Thus, investigating micro stores focuses attention on their negative role in the labor market that may prevent the achievement of such goals, especially given that they are the dominant business model in the Saudi economy.

Fourth, integrating theories of labor market segmentation into the analysis of the labor market allowed us to evaluate different public policies, such as spending on education and training, subsidies for the private sector, and localization policies. For example, according to the dual labor market theory, the return on human capital in the secondary labor market is less than that in the primary labor market (Doeringer & Piore, 1970; Osterman, 1975; Piore, 1968).

Fifth, technological progress has facilitated the shift toward capital-intensive production and automation, which may further increase unemployment rates. In addition, although growing international trade opens up different opportunities for both producers and consumers, it may also increase unemployment even further when firms move or outsource operations to other countries.

Finally, the Saudi government appears to be aware of its inability to continue as the first-and-last-resort employer for the increasing local labor force. Therefore, in recent years, the government has sought to increase Saudi employment in the private sector by enforcing localization policies to replace foreign workers with Saudi workers and by introducing an expatriate levy. However, localization policies could have counterproductive effects, especially

in the long run. Moreover, by using the nested CES production functions, we can predict the impact of increasing the cost of foreign workers in Saudi employment in both big and micro establishments.

In short, the market-distorting factors are important in the analysis of the Saudi labor market because they help explain at least seven central issues: (i) the high and persistent unemployment rate, (ii) the low labor participation rate among Saudi men and women or the slow process of integrating local labor into the private sector, (iii) the low human capital investment in specific skills through on-job-training programs such as internship and apprenticeship, (iv) the poor economic performance of the private sector and why it is susceptible to the localization policies and any changes in government programs (subsidies and purchases), (v) underemployment and low labor productivity in the public sector and its low resilience in the face of economic shocks resulting from fluctuations of oil prices, (vi) high dependence rates on foreign labor in all industries, and (vii) the slow process of economic transformation and diversification.

6.6 Limitations of the study

The limitations of this study can be summarized as follows. First, the research is based on the official classifications of big industries (i.e., 20 employees or more), which may not best represent big industries. Hence, it is suggested to use, if possible, at least 100 employees as a definition of a big industry when the elasticity of substitution is being estimated in big industries. Second, it would be instructive to estimate the nested CES production function using the direct (economic) method when disaggregated data of wage levels and prices of different inputs are available. Then, the first-order conditions could be used in the estimation. Finally, because of the potential problem of multicollinearity between the local-labor (M_L) and foreign-labor (M_F) variables, it would be advisable to drop the local-labor variable and use the foreign-labor variable with capital to estimate the elasticity of substitution between the two variables. This

is because the local laborers in micro stores are reported as workers when they are, in fact, owners. However, this approach has a drawback because it would not allow for the two types of elasticity of substitution (direct and partial). In other words, only the general elasticity of substitution would be estimated.

6.7 Recommendations

Recommendations for researchers and policymakers can be summarized as follows. First, researchers could use the same or similar theoretical paradigm to investigate whether the elasticity of substitution varies among different industry sizes, instead of relying on the general estimation of the elasticity of substitution using aggregate data. The general estimation may not necessarily provide a better understanding of the substitution relationship between Saudi and non-Saudi labor.

Second, Lindbeck and Snower (2001) mention the “insider versus outsider” notion. This concept has important implications for the Saudi labor market and could be applied to the market to explain the labor relations (and possibly tensions) between insider employees (foreign labor) and outsider potential employees (Saudi labor). This model is expected to have useful implications for the Saudi market and to create a better understanding of labor relations in different industries or organizations.

Third, Lucas (1978) developed a model to explain the relationship between managers’ “talent” distribution and firms’ size distribution. This model suggests that the number of managers increases when the average wage level decreases, as in recession times. Thus, the number of firms increases (i.e., the distribution of small firms increases) as the ratio of employees to managers decreases. This model can be used to explain the increasing number of micro stores or Saudi businesses in general. This is for several reasons: (i) the low wage levels in the private sector on the one hand, and the restrained wages in the public sector on

the other, (ii) the low start-up costs of micro stores and other small businesses, and (iii) the accessibility of cheap foreign labor.

Fourth, policymakers and researchers should not only focus on how to implement localization policies to increase Saudi employment or on how to make national labor more competitive to increase chances of recruitment (labor supply). They should also focus on supporting the private sector to create meaningful jobs as well as recruiting and training national labor (labor demand), which would enhance voluntary transactions between employers and national labor.

Fifth, the high distribution of small establishments in the Saudi market constrains the recruitment and training of Saudi labor, particularly of women. Because Saudi women usually prefer to work in Islamically or socially accepted places, compliance with those requirements is costly for small businesses because of the low capacity for scale economies. Therefore, this research suggests that these costs could be mitigated by providing "female-only, shared workplaces", then renting these workplaces to different establishments. This would encourage participation and recruitment of Saudi women while meeting the needs of small and financially constrained businesses. As another advantage of this proposal, other supplementary services, such as childcare and transportation, could be provided effectively at lower average costs by such workplaces. Instead of providing grants for each business individually to establish female-only workplaces, the government could fund establishing those workplaces and earn profits. The government could also facilitate joint venture or cooperative funding among businesses to establish such new shared workplaces.

Sixth, the nested production function with three inputs of production has three possible structures. This thesis has estimated only one structure, which is nesting Saudi labor with capital, then these two inputs as a composite input with foreign labor. There are two other structures that have not been estimated: (i)

nesting Saudi labor with foreign labor, then these two inputs as a composite input with capital, and (ii) nesting foreign labor with capital, then these two inputs as a composite input with Saudi labor. Further research can be done to estimate the elasticity of substitution by investigating the latter structures.

Seventh, after identifying the three market-distorting factors, the thesis proposes two corrective measures to reverse or mitigate the negative effects of those factors. First, it is suggested that the sponsorship system be replaced by a central recruitment agency. Second, it is suggested that micro stores be re-zoned or restricted within self-contained centers. Consortiums between the government and private investors could be created to establish TDCs to develop certain industries and product markets while recruiting and training national labor, as explained extensively in section 3.4.

References

100 billion volumes of commercial cover-up in Saudi Arabia managed by expatriates .. and 6 measures to combat it. (2017, October 14). *Maaal*. https://www.maaal.com/archives/20171014/97841

Abdalla, I. M., Al-Waqfi, M. A., Harb, N., Hijazi, R. H., & Zubaidi, T. (2010). Labour Policy and Determinants of Employment and Wages in a Developing Economy with Labour Shortage. *Labour*, *24*(2), 163-177. https://doi.org/10.1111/j.1467-9914.2010.00474.x

Abdullah, I., Al-Sayah, T., & Al-Muhanna, H. (2013, January 8). Saudi spends more than 1 trillion on education in ten years. *Aleqtisadiah Newspaper,*. http://www.aleqt.com/2013/01/08/article_723145.html

Abusaaq, H. I. (2015). *Population Aging In Saudi Arabia* [SAMA Working Paper: WP/15/2]. Saudi Arabian Monetary Agency. https://www.sama.gov.sa/en-US/EconomicResearch/Pages/WorkingPapers.aspx

Ahmed, S. F. (1993). Saudization of the Saudi labor market: a field economic study [سعودة سوق العمل السعودي، المؤلف: أحمد، سيد فتحي]. *Institute of Diplomatic Studies, Ministry of Foreign Affairs, KSA*, *7*, 7-71. https://search.mandumah.com/Record/177933

Al-Asfour, A., & Khan, S. A. (2014). Workforce localization in the Kingdom of Saudi Arabia: issues and challenges. *Human Resource Development International*, *17*(2), 243-253. https://doi.org/10.1080/13678868.2013.836783

Al-Asmari, M. (2008). Saudi labor force: Challenges and ambitions. *JKAU: Arts & Humanities*, *16*(2), 19-59.

Al-Azzaz, A. S., & Yousef, A. A. (1999). A Proposed Approach to Study the Problem of Employing National Labor in the Kingdom of Saudi Arabia [منهج مقترح لدراسة مشكلة توظيف العمالة الوطنية بالمملكة العربية السعودية، المؤلف: العزاز، عبدالله و يوسف، السيد عبدالحليم]. *Public Administration Journal, the Institute of Public Administration*, *38*(4), 665-702.

Al-Buraik, M. A. (1988). *An economic investigation of foreign labor composition in Saudi Arabia's labor market* [Ph.D., West Virginia University]. USA.

Al-Dosary, A. S. (1991). *Towards the reduction of foreign workers in Saudi Arabia* [Ph.D., The University of Michigan]. USA.

Al-Dosary, A. S. (2004). HRD or manpower policy? Options for government intervention in the local labor market that depends upon a foreign labor force: the Saudi Arabian perspective. *Human Resource Development International*, *7*(1), 123-135. https://doi.org/10.1080/1367886022000041967

Al-Dosary, A. S., & Rahman, S. M. (2005). Saudization (Localization)–A critical review. *Human Resource Development International*, *8*(4), 495-502. https://doi.org/10.1080/13678860500289534

Al-Filali, I. Y., & Gallarotti, G. M. (2012). Smart development: Saudi Arabia's quest for a knowledge economy. *International Studies*, *49*(1-2), 47-76. https://doi.org/10.1177/0020881713504673

Al-Ghaith, M., & Al-Maashoug, M. (1996). Employing the national labor force in the private sector: Obstacles and solutions [توظيف العمالة المواطنة في القطاع الخاص: المعوقات ومداخل الحلول، المؤلف: الغيث، محمد و المعشوق، منصور]. *Institute of Public Administration, Riyadh, Saudi Arabia*.

Al-Hajjar, B., & Presley, J. R. (1992). Constraints on development: small businesses in Saudi Arabia. *Middle Eastern Studies*, *28*(2), 333-351. https://doi.org/10.1080/00263209208700904

Al-Juhani, E. A. (1996). Increased Productivity and Wage Policy and their Ability to Influence the Process of National Employment in the Saudi Private Sector [زيادة الإنتاجية وسياسة الأجور ومدى قدرتهما على التأثير في عملية إحلال العمالة المواطنة في القطاع الخاص السعودي، المؤلف: الجهني، عيد]. *Public Administration Journal, the Institute of Public Administration*, *35*(4), 557-574.

Al-Maliki, J. (2015, February 25). 20 billion riyals annually for 150 thousand scholarship awaiting «unemployment». *Al-Madina Newspaper,*. https://bit.ly/38EayxB

Al-Mudayfer, T. (2005, Oct 20). Riyadh is a shocking congestion. *Al-Riyadh Newspaper*. https://www.alriyadh.com/101979

Al-Mutairi, Y. (2015, December 2). A memorable speech by the deceased “Ghazi Al-Qusaibi” blames businessmen for not appreciating Saudi youth. *Almnatiq Newspaper*. https://almnatiq.net/162489/

https://www.youtube.com/watch?v=eJIvoAm1DrA&feature=emb_title

Al-Qais, I. (2015, Nov. 21). Al-Batha ... a real obstacle to the "financial and business capital". *Al-Riyadh Newspaper*. https://www.alriyadh.com/1102537

Al-Rayes, R. M., & Al-Abed, M. R. (2017). *Explanation of the Articles of the Saudi Labor Law: In accordance with the amendments issued by the Royal Decree No (M/46) dated 5/6/1436 AH* [شرح مواد نظام العمل السعودي وفقا للتعديلات الصادرة بالمرسوم الملكي رقم (م/٤٦) وتاريخ ٥ /٦ /١٤٣٦ هـ، المؤلف: الريس، رزق و العبد، رضا محمود]. Riyadh: Alshegrey.

Al-Sudani, A. A., & Abdulkheer, A. Y. (2001). The future of Saudization and growth of labor productivity [مستقبل السعودة ونمو إنتاجية العمل، المؤلف: السوداني، عبدالعزيز علي و عبدالخير، أحمد يوسف]. *Public Administration Journal, the Institute of Public Administration*, *41*(1), 1-50.

Al-Thaqafi, S. A. (2000). The social, economic and security dimensions of increasing expatriate labor and shortage of the national labore in the private sector [الأبعاد الاجتماعية و الاقتصادية و الأمنية لزيادة العمالة الوافدة و نقص توظيف القوى العاملة الوطنية في القطاع الأهلي، المؤلف: الثقفي، سلطان]. *Public Administration Journal, the Institute of Public Administration*, *40*(3), 535-573.

Al Akayleh, F. (2016). Do Foreign Workers’ Positive Contributions to GDP Outweigh the Negative Effect of Their Outward Remittances on GDP? A Case Study of Saudi Arabia. *Advances in Economics and Business*, *4*(12), 657-666. https://doi.org/10.13189/aeb.2016.041204

Al Hajjar, B., & Presley, J. R. (1996). Small business in Saudi Arabia. In W. J. W. (eds) (Ed.), *Business and economic development in Saudi Arabia* (pp. 105-126). London: Palgrave Macmillan. https://doi.org/10.1007/978-1-349-25096-7_8

Al Salem, S. (2010, Jone 13). The crisis of parking spaces in front of apartments reflects the absence of "proper planning". *Al-Riyadh Newspaper*. https://www.alriyadh.com/534483

Al Sheikh, S. A. (2001). Demography movement in the Kingdom of Saudi Arabia and its effect on the labour market and economic growth. *The Saudi Economic Journal*, *Spring*(5), 68-73.

Al Zahrani, K. (2017, June 4). 0.8% is the percentage of Unemployed Saudi Engineers to non-Saudi Engineers. *Alwatan Newpaper,*. http://www.alwatan.com.sa/Local/News_Detail.aspx?ArticleID=305527&CategoryID=5

Aldukheil, A. (2013). *Saudi government revenues and expenditures: a financial crisis in the making*. Springer.

Alodadi, A., & Benhin, J. (2015). Long term economic growth in oil-rich Saudi Arabia: What is the role for non-oil sectors. *Economic Issues*, *20*(1), 109-129.

Alrasheedy, A. (2019). The Cost of Unemployment in Saudi Arabia. *International Journal of Economics and Finance*, *11*(11), 30-42. https://doi.org/10.5539/ijef.v11n11p30

Anderson, R. K., & Moroney, J. R. (1994). Substitution and Complementarity in C.E.S. Models. *Southern Economic Journal*, *60*(4), 886-895. https://doi.org/10.2307/1060427

Angrist, J. D., & Krueger, A. B. (1999). Chapter 23 - Empirical Strategies in Labor Economics. In O. C. Ashenfelter & D. Card (Eds.), *Handbook of Labor Economics* (Vol. 3, pp. 1277-1366). Elsevier. https://doi.org/10.1016/S1573-4463(99)03004-7

Arnott, R., & McMillen, D. (2006). *A companion to urban economics*. Oxford: Blackwell Publishing.

Arrow, K. J., Chenery, H. B., Minhas, B. S., & Solow, R. M. (1961). Capital-Labor Substitution and Economic Efficiency. *The Review of Economics and Statistics*, *43*(3), 225-250. https://doi.org/10.2307/1927286

Assaad, R. (2014). Making sense of Arab labor markets: the enduring legacy of dualism. *IZA Journal of Labor and Development*, *3*(1), 1-25. https://doi.org/10.1186/2193-9020-3-6

Assidmi, L. M., & Wolgamuth, E. (2017). Uncovering the dynamics of the Saudi youth unemployment crisis. *Systemic Practice and Action Research*, *30*(2), 173-186. https://doi.org/10.1007/s11213-016-9389-0

Baffoe-Bonnie, J. (2003). Distributional assumptions and a test of the dual labor market hypothesis. *Empirical Economics*, *28*, 461-478. https://doi.org/10.1007/s001810200141

Baqadir, A., Patrick, F., & Burns, G. (2011). Addressing the skills gap in Saudi Arabia: does vocational education address the needs of private sector employers? *Journal of Vocational Education & Training*, *63*(4), 551-561. https://doi.org/10.1080/13636820.2011.589533

Barron, J. M., Black, D. A., & Loewenstein, M. A. (1987). Employer Size: The Implications for Search, Training, Capital Investment, Starting Wages, and Wage Growth. *Journal of Labor Economics*, *5*(1), 76-89. http://www.jstor.org/stable/2535035

Becker, G. S. (1993). *Human capital: A theoretical and empirical analysis, with special reference to education* (3rd ed.). University of Chicago press.

Betcherman, G. (2013). *Labor market institutions: A review of the literature*. The World Bank.

Bhaskar, V., Manning, A., & To, T. (2002). Oligopsony and monopsonistic competition in labor markets. *Journal of Economic Perspectives*, *16*(2), 155-174. https://www.jstor.org/stable/2696501

Bin Obaid, A. S. (2003). Determinants of Employment in the Private Sector in the GCC Countries [محددات التوظيف في القطاع الخاص بدول مجلس التعاون الخليجي، المؤلف: ابن عبيد، أحمد]. *Kuwait University, Scientific Publishing Council*, *31*(1), 7-25. http://search.mandumah.com/Record/26642

Birks, J. S., & Sinclair, C. A. (1979). Some aspects of the labor market in the Middle East with special reference to the gulf states. *Journal of Developing Areas*, *13*(3), 301-318. https://www.jstor.org/stable/4190663

Black, D. A., Noel, B. J., & Wang, Z. (1999). On-the-Job Training, Establishment Size, and Firm Size: Evidence for Economies of Scale in the Production of Human Capital. *Southern Economic Journal*, *66*(1), 82-100. https://doi.org/10.2307/1060836

Borjas, G. J. (2003). The Labor Demand Curve Is Downward Sloping: Reexamining the Impact of Immigration on the Labor Market. *The Quarterly Journal of Economics*, *118*(4), 1335-1374. http://www.jstor.org/stable/25053941

Borjas, G. J. (2013). The analytics of the wage effect of immigration. *IZA Journal of Migration*, *2*(1), 1-25. https://doi.org/10.1186/2193-9039-2-22

Borjas, G. J. (2014). *Immigration economics*. Harvard University Press.

Boyer, G. R., & Smith, R. S. (2001). The development of the neoclassical tradition in labor economics. *ILR Review*, *54*(2), 199-223. https://doi.org/10.2307/2696007

Bulow, J. I., & Summers, L. H. (1986). A theory of dual labor markets with application to industrial policy, discrimination, and Keynesian unemployment. *Journal of Labor Economics*, *4*(3, Part 1), 376-414. https://www.jstor.org/stable/2535059

Cain, G. G. (1976). The Challenge of Segmented Labor Market Theories to Orthodox Theory: A Survey. *Journal of Economic Literature*, *14*(4), 1215-1257. https://www.jstor.org/stable/2722547

Card , D., & Krueger, A. B. (2017). Myth and Measurement and the Theory and Practice of Labor Economics. *ILR Review*, *70*(3), 826-831.

Card, D., & Lemieux, T. (2001). Can Falling Supply Explain the Rising Return to College for Younger Men? A Cohort-Based Analysis. *The Quarterly Journal of Economics*, *116*(2), 705-746. http://www.jstor.org/stable/2696477

Chaaban, J. (2009). Youth and Development in the Arab Countries: The Need for a Different Approach. *Middle Eastern Studies*, *45*(1), 33-55. https://www.jstor.org/stable/40262641

Chang, K. P. (1994). Capital-energy substitution and the multi-level CES production function. *Energy Economics*, *16*(1), 22-26. https://doi.org/10.1016/0140-9883(94)90014-0

Chirinko, R. S. (2008). [sigma]: The long and short of it [article]. *Journal of Macroeconomics*, *30*(2), 671-686. https://doi.org/10.1016/j.jmacro.2007.10.010

Citino, N. J. (2002). *From Arab Nationalism to OPEC: Eisenhower, King Sa'ūd, and the Making of US-Saudi Relations*. Houghton Mifflin Harcourt.

Cobb, C. W., & Douglas, P. H. (1928). A Theory of Production. *The American economic review*, *18*(1), 139-165. https://www.jstor.org/stable/1811556

Colander, D., Holt, R., & Rosser, B. (2004). The changing face of mainstream economics [article]. *Review of Political Economy*, *16*(4), 485-499. https://doi.org/10.1080/0953825042000256702

Demekas, D., & Kontolemis, Z. (2000). Government employment and wages and labor market performance. *Oxford Bulletin of Economics and Statistics*, *62*(3), 391-415. https://doi.org/10.1111/1468-0084.00177

Demekas, D. G. (1990). Labor market segmentation in a two-sector model of an open economy. *Staff Papers (International Monetary Fund)*, *37*(4), 849-864. https://doi.org/10.2307/3867231

Demsetz, H. (1973). Industry structure, market rivalry, and public policy. *Journal of Law and Economics*, *16*(1), 1-9. https://doi.org/10.1086/466752

Dequech, D. (2007). Neoclassical, Mainstream, Orthodox, and Heterodox Economics. *Journal of Post Keynesian Economics*, *30*(2), 279-302. http://www.jstor.org/stable/27746799

Dickens, W. T., & Lang, K. (1988). The Reemergence of Segmented Labor Market Theory. *The American economic review*, *78*(2), 129-134. https://www.jstor.org/stable/1818110

Dissou, Y., Karnizova, L., & Sun, Q. (2015). Industry-level Econometric Estimates of Energy-Capital-Labor Substitution with a Nested CES Production Function. *Atlantic Economic Journal*, *43*(1), 107-121. https://doi.org/10.1007/s11293-014-9443-1

Doeringer, P. B., & Piore, M. J. (1970). *Internal Labor Markets and Manpower Analysis*. Harvard University, Massachusetts Institution of Technology. https://eric.ed.gov/?id=ED048457

Edens, D. G., & Snavely, W. P. (1970). Planning for Economic Development in Saudi Arabia. *Middle East Journal*, *24*(1), 17-30. https://www.jstor.org/stable/4324550

Ehrenberg, R., & Schwarz, J. (1986). Chapter 22 Public-sector labor markets. In *Handbook of Labor Economics* (Vol. 2, pp. 1219-1260). Elsevier. https://doi.org/10.1016/S1573-4463(86)02012-6

Ehrenberg, R., & Smith, R. (2012). *Modern labor economics: theory and public policy*. Pearson Education.

Ethridge, D. E. (2004). *Research methodology in applied economics: organizing, planning, and conducting economic research*. Blackwell publishing Ames.

Farkas, G., & England, P. (Eds.). (1988). *Industries, Firms, and Jobs: sociological and economic approaches*. New York: Springer Science+Business Media.

Fasano, U., & Goyal, R. (2004). *Emerging strains in GCC labor markets*. International Monetary Fund.

Feess, E. (2012). Minimum wages in gulf states –an economic analysis. In S. Hertog (Ed.), *National employment, migration and education in the GCC* (pp. 7-64). Gerlach Press. https://doi.org/10.2307/j.ctt1s474nj.4

Feng, S., & Zhang, K. (2018). Fuel-factor nesting structures in CGE models of China. *Energy Economics*, *75*, 274-284. https://doi.org/10.1016/j.eneco.2018.08.030

Freeman, R. B. (1986). Unionism comes to the public sector. *Journal of Economic Literature*, *24*(1), 41-86. www.jstor.org/stable/2725787

Freeman, R. B. (1993). Labor Markets and Institutions in Economic Development. *The American economic review*, *83*(2), 403-408. https://www.jstor.org/stable/2117698

Freeman, R. B. (2005). Labour market institutions without blinders: The debate over flexibility and labour market performance. *International Economic Journal*, *19*(2), 129-145. https://doi.org/10.1080/10168730500080675

Gelb, A., Knight, J. B., & Sabot, R. H. (1991). Public sector employment, rent seeking and economic growth. *The Economic Journal*, *101*(408), 1186-1199. https://doi.org/10.2307/2234435

General Authority for Statistics. (2010a). *Enumerating establishments* https://www.stats.gov.sa/en/16

General Authority for Statistics. (2010b). *The general population and housing Census, 2010* https://www.stats.gov.sa/en/13

General Authority for Statistics. (2016). *the unemployment rate of foreign labor* https://www.stats.gov.sa/en/820

General Authority for Statistics. (2017a). *Annual economic establishments survey* https://www.stats.gov.sa/en/64

General Authority for Statistics. (2017b). *Average monthly wages per paid employee (15+) by sex, nationality and type of sector (SR)* https://www.stats.gov.sa/en/814

General Authority for Statistics. (2017c). *Gross domestic product* https://www.stats.gov.sa/en/823

General Authority for Statistics. (2017d). *Labor market, total labor force by sex, nationality, and education level* https://www.stats.gov.sa/en/814

General Authority for Statistics. (2017e). *Labor market, unemployment rate by sex, nationality and education level* https://www.stats.gov.sa/en/820

General Authority for Statistics. (2018). *Annual economic establishments survey* https://www.stats.gov.sa/ar/64

General Authority for Statistics. (2019a). *Labor force* https://www.stats.gov.sa/ar/816

General Authority for Statistics. (2019b). *Labor market, unemployment rate by sex, nationality and age group* https://www.stats.gov.sa/ar/814

General Authority for Statistics. (2019c). *Percentage distribution of Saudi unemployed Persons (15+) holders of secondary education or equivalent by sex and educational specialization* https://www.stats.gov.sa/ar/814

Greenland, S., Senn, S. J., Rothman, K. J., Carlin, J. B., Poole, C., Goodman, S. N., & Altman, D. G. (2016). Statistical tests, P values, confidence intervals, and power: a guide to misinterpretations. *European Journal of Epidemiology*, *31*(4), 337-350. www.jstor.org/stable/44851769

Haber, S., Cordes, J., & Barth, J. (1988). *Employment and Training Opportunities in Small and Large Firms*. U.S. Small Business Administration.

Harrison, B., & Sum, A. (1979). The Theory of "Dual" or Segmented Labor Markets. *Journal of Economic Issues*, *13*(3), 687-706. https://www.jstor.org/stable/4224841

Harry, W. (2007). Employment creation and localization: the crucial human resource issues for the GCC. *The International Journal of Human Resource Management*, *18*(1), 132-146. https://doi.org/10.1080/09585190601068508

Hazzazi, H. (2019, February 2). 900 Saudi dentists jobless when 9,000 expats work under MOH. *Saudi Gazette*. https://saudigazette.com.sa/article/558271

Henningsen, A., & Henningsen, G. (2011). *Econometric Estimation of the "Constant Elasticity of Substitution" Function in R: Package micEconCES* (FOI Working Paper 2011/9). University of Copenhagen, Department of Food and Resource Economics. https://orbit.dtu.dk/en/publications/econometric-estimation-of-the-constant-elasticity-of-substitution

Henningsen, A., & Henningsen, G. (2012). On estimation of the CES production function—Revisited. *Economics Letters*, *115*(1), 67-69. https://doi.org/10.1016/j.econlet.2011.12.007

Henningsen, A., & Henningsen, G. (2015). *Package 'micEconCES'*. http://www.micEcon.org

Henningsen, A., Henningsen, G., & van der Werf, E. (2019). Capital-labour-energy substitution in a nested CES framework: A replication and update of Kemfert (1998). *Energy Economics*, *82*, 16-25. https://doi.org/10.1016/j.eneco.2017.12.019

Herb, M. (Ed.). (2019). *The politics of rentier states in the Gulf*. The Project on Middle East Political Science (POMEPS). https://pomeps.org/pomeps-studies-33-the-politics-of-rentier-states-in-the-gulf.

Hertog, S. (2006). *Segmented clientelism: the politics of economic reform in Saudi Arabia* [Ph.D., University of Oxford]. UK.

Hertog, S. (2016, 2016-12-12 - 2016-12-13). *Challenges to the Saudi distributional state in the age of austerity.* Saudi Arabia: Domestic, Regional and International Challenges, Singapore. http://eprints.lse.ac.uk/id/eprint/68625

Hertog, S. (2018). *Can we Saudize the labor market without damaging the private sector?* King Faisal Center for Research and Islamic Studies. https://kfcris.com/pdf/8fd49f41c540c860a49bcd8d4f2966a75be162612cffa.pdf

Heun, M. K., Santos, J., Brockway, P. E., Pruim, R., Domingos, T., & Sakai, M. (2017). From theory to econometrics to energy policy: Cautionary tales for policymaking using aggregate production functions. *Energies*, *10*(2), Article 203. https://doi.org/10.3390/en10020203

International Monetary Fund. (1997). *Financial systems and labor markets in the Gulf Cooperation Council countries.* International Monetary Fund. https://doi.org/10.5089/9781557757128.071

International Monetary Fund. (2018). *Saudi Arabia, selected issues* (IMF Country Report, Issue. https://www.imf.org/en/Publications/CR/Issues/2018/08/24/Saudi-Arabia-Selected-Issues-46197

Kapiszewski, A. (2000). Population, labour and education dilemmas facing GCC states at the turn of the century. Cross-Roads of the New Millennium: Proceedings of theTechnological Education and National Development (TEND) 2nd, April 8-10, 2000,, Abu Dhabi, United Arab Emirates.

Kayed, R. N., & Kabir Hassan, M. (2011). Saudi Arabia's economic development: entrepreneurship as a strategy. *International Journal of Islamic and Middle Eastern Finance and Management*, *4*(1), 52-73. https://doi.org/10.1108/17538391111122203

Kerr, C. (1950). Labor Markets: Their Character and Consequences. *The American economic review*, *40*(2), 278-291. https://www.jstor.org/stable/1818048

Khan, A. H. (1989). The Two-level CES Production Function for the Manufacturing Sector of Pakistan. *The Pakistan Development Review*, *28*(1), 1-12. https://www.jstor.org/stable/41259209

Koesler, S., & Schymura, M. (2015). Substitution Elasticities In A Constant Elasticity Of Substitution Framework - Empirical Estimates Using Nonlinear Least Squares [article]. *Economic Systems Research*(1), 101. https://doi.org/10.1080/09535314.2014.926266

Krueger, A. B., & Summers, L. H. (1986). *Reflections on the inter-industry wage structure* (NBER Working Paper No. 1968). National Bureau of Economic Research. http://www.nber.org/papers/w1968.pdf

KUCUK, M., KUMBAROGLU, G., & SARICA, K. (2018). Substitution elasticities in an energy-augmented CES production function: An empirical analysis for Turkey. *Journal of Economics and Political Economy*, *5*(2), 234-249.

Kumar, T. K., & Gapinski, J. H. (1974a). Nonlinear Estimation of the CES Production Function: Sampling Distributions and Tests in Small Samples. *Southern Economic Journal*, *41*(2), 258-266. https://doi.org/10.2307/1056737

Kumar, T. K., & Gapinski, J. H. (1974b). Nonlinear Estimation of the CES Production Parameters: A Monte Carlo Study. *The Review of Economics and Statistics*, *56*(4), 563-567. https://doi.org/10.2307/1924476

Kutner, M. H., Nachtsheim, C. J., Neter, J., & Li, W. (2005). *Applied linear statistical models* (5 ed.). McGraw-Hill Irwin New York.

Leber, A. (2019). Resisting rentierism: Labor market reforms in Saudi Arabia. In M. Herb (Ed.), *The Politics of Rentier States in the Gulf* (pp. 34-39). The Project on Middle East Political Science (POMEPS). https://pomeps.org/pomeps-studies-33-the-politics-of-rentier-states-in-the-gulf

Leontaridi, M. R. (1998). Segmented Labour Markets: Theory and Evidence. *Journal of Economic Surveys*, *12*(1), 63-101. https://doi.org/10.1111/1467-6419.00048

Leontief, W. (1986). *Input-output economics, edited by Wassily Leontief* (2nd ed. ed.). ProQuest Ebook Central. https://ebookcentral.proquest.com/lib/vu/detail.action?docID=4701165

Lever, H. M. (1996). Firm Size and Employment Determination in Dutch Manufacturing Industries. *Small Business Economics*, *8*(5), 389-396. https://www.jstor.org/stable/40228689

Lewis, E., & Peri, G. (2015). Chapter 10 - Immigration and the Economy of Cities and Regions. In G. Duranton, J. V. Henderson, & W. C. Strange (Eds.), *Handbook of Regional and Urban Economics* (Vol. 5, pp. 625-685). Elsevier. https://doi.org/https://doi.org/10.1016/B978-0-444-59517-1.00010-6

Lindbeck, A., & Snower, D. J. (2001). Insiders versus Outsiders. *The Journal of Economic Perspectives*, *15*(1), 165-188. https://www.jstor.org/stable/2696546

Lipczynski, J., Wilson, J., & Goddard, J. (2005). *Industrial organization: competition, strategy, policy* (2nd ed.). Prentiee Hall.

Looney, R. E. (1988). The impact of technology transfer on the structure of the Saudi Arabian labor force. *Journal of Economic Issues*, *22*(2), 485-492. https://doi.org/10.1080/00213624.1988.11504779

Looney, R. E. (1991). Patterns of human resource development in Saudi Arabia. *Middle Eastern Studies*, *27*(4), 668-678. https://doi.org/10.1080/00263209108700884

Loveridge, R., & Mok, A. L. (1979). *Theories of labour market segmentation: a critique* (1st ed.). Springer Science & Business Media.

Lucas, R. E. (1978). On the size distribution of business firms. *The Bell Journal of Economics*, *9*(2), 508-523. https://doi.org/10.2307/3003596

Madhi, S. T., & Barrientos, A. (2003). Saudisation and employment in Saudi Arabia. *Career Development International*, *8*(2), 70-77. https://doi.org/10.1108/13620430310465471

Mahdi, S. (2000). *Saudization and structure of the labour market in Saudi Arabia* [Economics Working Paper No. 24]. The University of Hertfordshire.

Manning, A. (2011). Chapter 11 - Imperfect Competition in the Labor Market. In D. Card & O. Ashenfelter (Eds.), *Handbook of Labor Economics* (Vol. 4, pp. 973-1041). Elsevier. https://doi.org/https://doi.org/10.1016/S0169-7218(11)02409-9

Marshall, A. (2013). *Principles of economics*. ProQuest Ebook Central. https://ebookcentral.proquest.com/lib/vu/detail.action?docID=1609118

Mashaal, A. A. (2013). Estimating the Demand for Foreign Labor in the Private Sector and the Possibility of Replacing Them with Saudi Labor, Using Johansen Cointegration and Error Correction Model [تقدير الطلب على العمالة الوافدة في القطاع الخاص وإمكانية إحلالها بالعمالة الوطنية في المملكة العربية السعودية باستخدام التكامل المشترك وتصحيح الخطأ، المؤلف: مشعل، أحمد عبداللطيف]. *Public Administration Journal, the Institute of Public Administration*, *53*(3), 703-737.

McDonald, I. M., & Solow, R. M. (1985). Wages and employment in a segmented labor market. *The Quarterly Journal of Economics*, *100*(4), 1115-1141. https://doi.org/10.2307/1885677

Mellahi, K. (2007). The effect of regulations on HRM: private sector firms in Saudi Arabia. *The International Journal of Human Resource Management*, *18*(1), 85-99. https://doi.org/10.1080/09585190601068359

'Ministry of Commerce': The size of the commercial concealment in Saudi Arabia reaches 400 billion riyals annually. (2019, February 12). *Argaam*. http://bit.ly/2tqRTBX

The [Saudi] Labor Law amended by Royal Decree No. (M/134), dated 27/11/1440 H. (29/7/2019), (2019). https://mlsd.gov.sa/en/policies/labor-law

Ministry of Municipal and Rural Affairs. (2018). *Shop licenses*. Retrieved August 8, 2018 from https://www.momra.gov.sa/GeneralServ/statistic.aspx

Mishrif, A., & Alabduljabbar, A. (2018). Quality of Education and Labour Market in Saudi Arabia. In A. Mishrif & Y. Al Balushi (Eds.), *Economic Diversification in the Gulf Region, V. I: The private sector as an engine of growth* (pp. 97-115). Singapore: Springer Singapore. https://doi.org/10.1007/978-981-10-5783-0_5

Mizon, G. E. (1977). Inferential Procedures in Nonlinear Models: An Application in a UK Industrial Cross Section Study of Factor Substitution and Returns to Scale. *Econometrica*, *45*(5), 1221-1242. https://doi.org/10.2307/1914069

Müller, T. (2003). Migration policy in a small open economy with a dual labor market. *Review of International Economics*, *11*(1), 130-143. https://doi.org/10.1111/1467-9396.00373

National Gas & Industrialization Company. (2020). *About Us*. Retrieved March 25, 2020 from http://www.gasco.com.sa/pages/view/about

Niblock, T., & Malik, M. (2007). *The political economy of Saudi Arabia*. London: Routledge. https://doi.org/10.4324/9780203946190

Nicholson, W., & Snyder, C. (2008). *Microeconomic theory: basic principles and extensions* (tenth ed.). S.-W. Cengage Learning.

Odrowąż-Coates, A. (2015). Cultural conditioning of the labour market in Saudi Arabia. *Politeja*, *12*(31/2), 185-198. https://www.jstor.org/stable/24919783

Oi, W. Y. (1983). The fixed employment costs of specialized labor. In J. E. Triplett (Ed.), *The measurement of labor cost* (pp. 63-122). Chicago: University of Chicago Press. http://www.nber.org/chapters/c7374

Oi, W. Y. (1990). Employment Relations in Dual Labor Markets ("It's Nice Work If You Can Get It"). *Journal of Labor Economics*, *8*(1), S124-S149. https://www.jstor.org/stable/2535210

Oi, W. Y., & Idson, T. L. (1999). Chapter 33 Firm size and wages [Review Article]. *Handbook of Labor Economics*, *3*(Part B), 2165-2214. https://doi.org/10.1016/S1573-4463(99)30019-5

Oliver, P. (2010). *The Student's Guide to Research Ethics* (second ed.). McGraw-Hill Education. http://ebookcentral.proquest.com/lib/vu/detail.action?docID=557103

Osterman, P. (1975). An Empirical Study of Labor Market Segmentation. *ILR Review*, *28*(4), 508-523. https://doi.org/10.1177/001979397502800402

Peck, J. (2019). *What Enables Firms to Hire Women? Overcoming Integration Costs in Saudi Arabia* (Commentaries, Issue. https://www.kfcris.com/en/view/post/252

Peck, J. R. (2017). Can hiring quotas work? The effect of the nitaqat program on the Saudi private sector. *American Economic Journal: Economic Policy*, *9*(2), 316-347. https://doi.org/10.1257/pol.20150271

Pereira, C. M. (2003). *Empirical Essays on the Elasticity of Substitution, Technical Change, and Economic Growth* [Ph.D., North Carolina State University]. USA.

Piore, M. J. (1968). On-the-job training and adjustment to technological change. *Journal of Human Resources*, *3*(4), 434-449. https://doi.org/10.2307/144796

Prywes, M. (1986). A nested CES approach to capital-energy substitution. *Energy Economics*, *8*(1), 22-28. https://doi.org/10.1016/0140-9883(86)90022-8

Ramady, M. A. (2010). *The Saudi Arabian economy: policies, achievements, and challenges* (second ed.). Springer Science & Business Media. https://doi.org/10.1007/978-1-4419-5987-4

Ramady, M. A. (2013). Gulf unemployment and government policies: prospects for the Saudi labour quota or Nitaqat system. *International Journal of Economics and Business Research*, *5*(4), 476-498. https://doi.org/10.1504/IJEBR.2013.054266

Ramady, M. A. (2014). *Political, economic and financial country risk, analysis of Gulf Cooperation Council*. Springer.

Rebitzer, J. B. (1993). Radical Political Economy and the Economics of Labor Markets. *Journal of Economic Literature*, *31*(3), 1394-1434. https://www.jstor.org/stable/2728245

Rihani, A. (1954). *The Modern History of Najd* [تاريخ نجد الحديث وملحقاته، المؤلف: الريحاني، أمين]. Beirut: Dar Sader Publishers.

Sadi, M. A., & Henderson, J. C. (2005). Local versus foreign workers in the hospitality and tourism industry: A Saudi Arabian perspective. *Cornell hotel and restaurant administration quarterly*, *46*(2), 247-257. https://doi.org/10.1177/0010880404273889

Sadi, M. A., & Henderson, J. C. (2010). Towards job localization in Saudi Arabia: Drivers and barriers within the services industry. *Journal of Immigrant & Refugee Studies*, *8*(2), 121-134.

Samuelson, P. A., & Nordhaus, W. D. (2010). *Economics* (19th ed.). McGraw-Hill Irwin.

Sato, K. (1967). A Two-Level Constant-Elasticity-of-Substitution Production Function. *The Review of Economic Studies*, *34*(2), 201-218. https://doi.org/10.2307/2296809

Saudi Arabian Monetary Authority. (2016). *Annals of Statistics*

Saudi Arabian Monetary Authority. (2017). *Annals of Statistics* http://www.sama.gov.sa/en-US/EconomicReports/Pages/YearlyStatistics.aspx

Saudi Arabian Monetary Authority. (2018). *Annals of Statistics* http://www.sama.gov.sa/en-US/EconomicReports/Pages/YearlyStatistics.aspx

Saudi Electricity Company. (2020). *Establishment and Merger and Structuring*. Retrieved March 26, 2020 from https://www.se.com.sa/en-us/Pages/aboutus.aspx

Saudi Press Agency. (2020, 1441/7/1 - 2020/02/25). *Royal Orders Issued* https://www.spa.gov.sa/viewstory.php?lang=en&newsid=2039164

Saudi Stock Exchange. (2019). *Company Information*. https://www.tadawul.com.sa/wps/portal/tadawul/home?locale=en

Saudi Vision 2030. (2016). *Saudi Vision 2030*. https://www.vision2030.gov.sa/en/node

Schmidt, C. M., & Zimmermann, K. F. (1991). Work Characteristics, Firm Size and Wages. *The Review of Economics and Statistics*, *73*(4), 705-710. https://doi.org/10.2307/2109410

Shaffer, S. (2009). Industrial structure and economic stability. *Applied Economics Letters*, *16*(6), 549-555. https://doi.org/10.1080/13504850701206536

Shalit, S. S., & Sankar, U. (1977). The measurement of firm size. *The Review of Economics and Statistics*, *59*(3), 290-298. https://doi.org/10.2307/1925047

Shapiro, C., & Stiglitz, J. E. (1984). Equilibrium Unemployment as a Worker Discipline Device. *The American economic review*, *74*(3), 433. https://www.jstor.org/stable/1804018

Shen, K., & Whalley, J. (2013). *Capital-Labor-Energy Substitution in Nested CES Production Functions for China* (NBER Working Paper No. 19104). National Bureau of Economic Research, Inc. http://doi.org/10.3386/w19104

Sherbiny, N. A. (1984a). DEVELOPMENT AND THE LABOR MARKET IN SAUDI ARABIA. *Journal of Arab Affairs*, *3*(1), [47].

Sherbiny, N. A. (1984b). Expatriate labor flows to the Arab oil countries in the 1980s. *Middle East Journal*, *38*(4), 643-667. https://www.jstor.org/stable/4326921

Sirageldin, I. A., Sherbiny, N. A., & Serageldin, M. I. (1984). *Saudis in transition: the challenges of a changing labor market*. Washington, D.C.: World Bank Group. http://documents.worldbank.org/curated/en/937661468777907398/Saudis-in-transition-the-challenges-of-a-changing-labor-market

Smith, R., & Zoega, G. (2009). Keynes, investment, unemployment and expectations [Article]. *International Review of Applied Economics*, *23*(4), 427-444. https://doi.org/10.1080/02692170902954767

Solow, R. M. (1980). On Theories of Unemployment. *The American economic review*, *70*(1), 1-11. https://www.jstor.org/stable/1814733

Spiess, A. (2010). Demographic transitions and imbalances in the GCC: security risks, constraints and policy challenges. The 2010 Exeter Gulf Studies Conference: Proceedings of The 21st-Century Gulf: The Challenge of Identity, 30 June - 3 July, 2010,

Stern, D. I. (2011). Elasticities of substitution and complementarity. *Journal of Productivity Analysis*, *36*(1), 79. https://www.jstor.org/stable/23883950

Stevens, P. (1986). *The impact of oil on the role of the state in economic development-A case study of the Arab world*. Surrey Energy Economics Centre, School of Economics, University of Surrey. https://ideas.repec.org/p/sur/seedps/29.html

Taubman, P., & Wachter, M. L. (1986). Chapter 21 Segmented labor markets. In *Handbook of Labor Economics* (Vol. 2, pp. 1183-1217). Elsevier. https://doi.org/10.1016/S1573-4463(86)02011-4

U.S. Energy Information Administration. (2020). *Saudi Arabia major oil and gas infrastructure*. Retrieved December 8, 2020 from https://www.eia.gov/international/analysis/country/SAU

Van der Werf, E. (2008). Production functions for climate policy modeling: An empirical analysis. *Energy Economics*, *30*(6), 2964-2979. https://doi.org/10.1016/j.eneco.2008.05.008

Vassiliev, A. (2000). *The History of Saudi Arabia*. London: Saqi Books.

Wackerly, D. D., Mendenhall, W., & Scheaffer, R. L. (2008). *Mathematical Statistics with Applications* (7th ed.). Thomson Learning, Inc.

Waterson, M. (1987). Recent developments in the theory of natural monopoly. *Journal of Economic Surveys*, *1*(1-2), 59-80. https://doi.org/10.1111/j.1467-6419.1987.tb00024.x

Wilson, R., Al Salamah, A., Malik, M., & Al Rajhi, A. (2012). *Economic development in Saudi Arabia*. London: Routledge. https://doi.org/10.4324/9780203037577

Wooldridge, J. M. (2013). *Introductory Econometrics: A Modern Approach* (5th ed.). South-Western Cengage Learning.

World Bank. (2018). *Personal remittances, paid (current US$)* https://data.worldbank.org/indicator/BM.TRF.PWKR.CD.DT?locations=SA&year_high_desc=true

Wright, J., al-Hajjar, B., & Presley, J. R. (1996). Attitudes, culture and capital distribution in Saudi Arabia. In W. J. W. (eds) (Ed.), *Business and economic development in Saudi Arabia* (pp. 127-139). London: Palgrave Macmillan. https://doi.org/10.1007/978-1-349-25096-7_9

Wright, J. W., Khashoggi, H., & Vaughn, C. (1996). Labour constraints on Saudi business development. In W. J. W. (eds) (Ed.), *Business and economic development in Saudi Arabia* (pp. 97-104). London: Palgrave Macmillan. https://doi.org/10.1007/978-1-349-25096-7_7

You, J.-I. (1995). Small firms in economic theory. *Cambridge Journal of Economics*, *19*(3), 441-462. https://doi.org/10.1093/oxfordjournals.cje.a035323

Zamberi Ahmad, S. (2012). Micro, small and medium-sized enterprises development in the Kingdom of Saudi Arabia. *World Journal of Entrepreneurship, Management and Sustainable Development*, *8*(4), 217-232. https://doi.org/10.1108/20425961211276606

Appendices

A1. Some Images of Micro Stores in Saudi Arabia

Appendix 1: Some Images of Micro Stores in Saudi Arabia

(a) (b) (c)

Note: Pictures of some of the micro stores that exist all along the city grid on the so-called commercial streets. The excessive number of micro stores and their unattractive features are clearly visible. Even recently established stores lack supplementary services, such as sufficient parking space and other amenities. Typically, these new micro stores have Saudi sponsors who rent them out, obtain foreign workers' visas, and then sublease the stores back to the foreign workers in a process known as “commercial concealment,” leading to the continuous reliance on foreign workers.

Sources:

(a) (Al-Qais, 2015), (b) (Al Salem, 2010), (c) (Al-Mudayfer, 2005), and all other pictures are taken by the researcher.

A2. Saudi Market Structure

Appendix 2: Percentage Distributions of Industries by Establishment Size and Employment Nationality Ranked by Industry Size							
Economic Activity	Percentage of Industry Size	Percentage of Micro Est.	Percentage of Small Est.	Percentage of Large Est.	Percentage of Saudi Employees	Percentage of Non-Saudi Employees	Percentages of Total Employees
Wholesale and retail trade; repair of motor vehicles	47.62	88.45	10.49	1.05	20.73	79.27	26.87
Manufacturing	10.94	79.81	15.71	4.47	20.08	79.92	16.04
Accommodation and food service activities	10.55	77.16	20.55	2.29	14.86	85.14	8.31
Agriculture, forestry and fishing	9.95	94.57	4.99	0.44	14.48	85.52	5.68
Other service activities	7.08	93.28	6.29	0.44	20.28	79.72	3.05
Construction	3.28	54.07	33.78	12.15	11.47	88.53	16.34
Real estate activities	2.95	91.96	6.90	1.13	48.35	51.65	1.72
Administrative and support service activities	1.89	76.66	18.61	4.72	26.42	73.58	3.89
Transportation and storage	1.63	62.10	29.51	8.39	30.73	69.27	3.88
Professional, scientific and technical activities	1.16	68.16	26.99	4.85	26.66	73.34	1.57
Education	0.82	25.14	44.91	29.95	47.25	52.75	2.74
Financial and insurance	0.56	40.69	48.70	10.61	68.28	31.72	1.55
Human health and social work activities	0.49	24.32	44.49	31.19	27.30	72.70	2.77
Information and communication	0.49	71.48	21.19	7.33	67.21	32.79	1.42
Water supply; sewerage, waste remediation	0.23	71.33	19.45	9.22	20.12	79.88	0.66
Arts, entertainment and recreation	0.23	64.93	27.78	7.29	16.03	83.97	0.50
Electricity, gas, steam and air conditioning supply	0.07	55.44	18.09	26.47	75.51	24.49	1.05
Mining and quarrying	0.06	10.73	44.53	44.74	65.27	34.73	1.95
Source: Enumerating Establishments (General Authority for Statistics, 2010a)							

A3. International Standard Industrial Classification (ISIC4)

Appendix 3: International Standard Industrial Classification of All Economic Activities (ISIC4)		
Industry	**Code**	**Economic Activity Description**
A. Agriculture, forestry and fishing	01	Crop and animal production, hunting and related service activities
	02	Forestry and woodcutting
	03	Fishing and aquaculture
B. Mining and quarrying	05	Mining of coal and lignite
	06	Extraction of crude oil and natural gas
	07	Mining of metal ores
	08	Other mining and quarrying activities
	09	Mining support services
C. Manufacturing	10	Manufacture of food products
	11	Manufacture of beverages
	12	Manufacture of tobacco products
	13	Manufacture of textiles
	14	Manufacture of wearing apparel
	15	Manufacture of leather products and related products
	16	Manufacture of wood and of products of wood and cork, except furniture
	17	Manufacture of paper and paper products
	18	Printing and reproduction of registered media materials
	19	Manufacture of coke and refined petroleum products
	20	Manufacture of chemicals and chemical products
	21	Manufacture of pharmaceutics and basic pharmaceutical products
	22	Manufacture of rubber and plastics products
	23	Manufacture of non-metallic products
	24	Manufacture of basic metals
	25	Manufacture of fabricated metal products, except machinery and
	26	Manufacture of computer, electronic and optical products
	27	Manufacture of electrical devices
	28	Manufacture of equipment and not elsewhere classified products
	29	Manufacture of motor vehicles, trailers and semi-trailers
	30	Manufacture of other transport equipment
	31	Furniture industry
	32	Other manufacturing industries
	33	Repair and installation of machinery and equipment
D. Electricity, gas, steam and air conditioning supply	35	Supplies of electricity, gas, steam and air conditioning

E. Water supply; sewerage, waste remediation	36	Water collection, treatment and supply
	37	Sewage
	38	Waste collection, treatment & disposal activities; materials recovery
	39	Remediation activities and other waste management services
F. Construction	41	Construction of buildings
	42	Civil engineering
	43	Specialized construction activities
G. Wholesale and retail trade; repair of motor vehicles	45	Wholesale and retail trade, and motor vehicles and motorcycle repair
	46	Wholesale trade, except of motor vehicles and motorcycles
	47	Retail trade, except of motor vehicles and motorcycles
H. Transportation and storage	49	Land transport and transport via pipelines
	50	Water transport
	51	Air transport
	52	Warehousing and support activities for transportation
	53	Post and courier activities via representatives
I. Accommodation and food service activities	55	Accommodation
	56	Food and beverage service activities
J. Information and communication	58	Publishing activities
	59	Motion picture, video & TV programme production, sound and music
	60	Programming and broadcasting activities
	61	Telecommunications
	62	Computer programming, consultancy and related activities
	63	Information service activities
K. Financial and insurance activities	64	Other financial service activities, except insurance and pension funding
	65	Insurance, reinsurance and pension funding
	66	Activities auxiliary to financial services and insurance activities
L. Real estate activities	68	Real estate activities
M. Professional, scientific and technical activities	69	Legal and accounting activities
	70	Activities of head offices; management consultancy activities
	71	Architectural and engineering activities; technical testing & analysis
	72	Scientific research and development
	73	Advertising and market research
	74	Other professional, scientific and technical activities
	75	Veterinary activities
N. Administrative and support service activities	77	Rental and leasing activities
	78	Employment activities

	79	Travel agencies, tour operators, reservation service and related activities
	80	Security and investigation activities
	81	Services to buildings and landscape activities
	82	Office administrative, office support and other business support activities
P. Education	85	Education
Q. Human health and social work activities	86	Human health activities
	87	Residential care activities with accommodation
	88	Social work activities without accommodation
R. Arts, entertainment and recreation	90	Arts, entertainment activities and recreation
	91	Libraries, archives, museums and other cultural activities
	93	Sports activities and amusement and recreation activities
S. Other service Organizations & bodies	94	Activities of membership organizations
	95	Repair of computers and personal and household goods
	96	Other personal service activities
Source: Establishments Economic Survey (General Authority for Statistics, 2017a)		