

Local Impacts of Natural Resource Booms and Busts



Gerhard Toews

St Antony's College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy in Economics

Trinity Term 2014

Local Impacts of Natural Resource Booms and Busts

Gerhard Toews

St Antony's College, University of Oxford

A thesis submitted for the degree of Doctor of Philosophy in Economics

TrinityTerm 2014

Abstract

This thesis consists of five stand-alone chapters empirically evaluating questions relating to the life cycle of natural resource extraction. We use three different data sets to shed light on the local impacts of natural resource booms and busts. In chapter 2 to 4 we use the household budget survey of Kazakhstan to explore the impacts of the oil boom on the local population. In the second chapter, we explore the distributional effects of the oil boom and show that average income increased and income inequality decreased. In the third chapter we study how the increase in average income was perceived by the local population and find that households' satisfaction with income *decreased*. In the fourth chapter we study how the boom affected households' expenditure and show that the likelihood that households pay tuition fees for tertiary education increased. In chapter 5, we explore the long-term impacts of a negative labour demand shock following the coal mine closures in the UK. To do this we construct a new data set containing the location of all active coal mines since 1981 and link it to the UK census. We find that the dramatic lay off of miners since 1981 was associated with a *persistent reduction* in female labour force participation in the affected districts. In chapter 6, we study the determinants of drilling costs and their impact on the real price of oil using a new global data set on the number of exploration wells drilled and costs of drilling. To do this, we propose a structural model of the *upstream sector* in the oil and gas industry. The model allows us to decompose the variation in the reduced form errors of the estimated VAR into three structural shocks, and estimate the dynamic responses of the variables in the system to these shocks. We confirm that the upstream sector of the oil and gas industry is subject to increasing costs. But we do *not* find that the real oil price is permanently affected by shocks to costs of drilling.

Word Count

The approximate number of words in this thesis is 53995 excluding the bibliography. This is calculated using the number of words on page 4, 327, multiplied by the number of pages, 165.

Declaration

I declare that this thesis represents my own work, and that none of it has already been accepted, or is currently being submitted, for any degree or diploma or certificate or other qualification in this University or elsewhere. Chapter 4 is a paper which has been jointly written with Alexander Libman. I have contributed the most work on this paper. Chapter 5 is an adjusted version of a jointly authored paper with Fernando Aragon and Juan Pablo Rud. I have also contributed the most work on this paper.

Acknowledgments

I would like to express my sincere gratitude to my supervisor Tony Venables. During the whole process he was very critical regarding the questions I wanted to ask in my research. Even though painful, this was an essential part of developing this thesis. I would also like to thank Celia Kingham who made me feel comfortable at the economics department.

I am grateful for several people supporting me remotely in these past years. A sincere thanks goes to Alexander Naumov who was extremely helpful in the process of my thesis. I would also like to thank Anton and Irina Epp, Eduard Giss, Arthur Lueck, Zeyyad Mandelinici, Waldemar and Lilia Muth, and Eugen Potorac for their support.

I am grateful to my parents, Valentina and Alfred Toews, who have been very supportive and motivating throughout the difficult times of writing this thesis.

Foremost, I am thankful to Martina Kirchberger who was of fundamental importance in making me stay in Oxford and finish my DPhil. I am also grateful for all the discussions which allowed me to gain a better understanding of my thesis and the world surrounding us.

Contents

1	Introduction	1
2	Distributional Effects of Natural Resource Booms: Evidence from Kazakhstan	3
2.1	Introduction	3
2.2	Oil Boom and Linkages to the Local Economy	6
2.3	Theoretical Framework	11
2.4	Data	16
2.5	Identification and Estimation Strategy	18
2.6	Results	22
2.7	Alternative Explanations and Concerns	25
2.8	Conclusion	27
2.9	Appendix	29
3	Inflated Expectations and Natural Resource Booms: Evidence from Kazakhstan	38
3.1	Introduction	38
3.2	Oil Boom and Conflicts	41
3.3	Theoretical Framework	44
3.4	Data	45
3.5	Identification and Estimation Strategy	48
3.6	Results	54
3.7	Alternative explanations	58
3.8	Conclusion	61
3.9	Appendix	62

4	Getting incentives right: Human Capital Investment and Natural Resource Booms	72
4.1	Introduction	72
4.2	Empirical Case: Kazakhstan	76
4.3	Theoretical Framework	79
4.4	Data	82
4.5	Identification and estimation	84
4.6	Results	89
4.7	Discussion	94
4.7.1	Is our result driven by the Engel curve effect?	94
4.7.2	Is our result driven by increased supply of education?	95
4.7.3	Is our result driven by educated immigrants?	99
4.7.4	Is our result driven by peer effects and educational signaling?	100
4.8	Conclusion	103
4.9	Appendix	105
5	Mine closures and female employment: Evidence from the UK	111
5.1	Introduction	111
5.2	Empirical Case: Coal Industry in the UK	113
5.3	Theoretical Framework	116
5.3.1	Baseline Model	116
5.3.2	Agglomerations	118
5.3.3	Non-Tradable Goods	119
5.3.4	Labour participation and unemployment	119
5.3.5	Heterogeneous labour	120
5.3.6	Empirical Predictions	120
5.4	Data	121
5.5	Identification and Estimation	124
5.6	Results	126
5.6.1	Substitution of Female Workers	129
5.6.2	Robustness	132

5.7	Conclusion	133
5.8	Appendix	135
6	Modelling Shocks in the Upstream Sector of the Oil and Gas Industry	139
6.1	Introduction	139
6.2	Upstream Sector in the Oil and Gas Industry	142
6.3	Theoretical Framework	145
6.4	Data	148
6.5	Identification and Estimation	152
6.6	Results	155
6.7	Conclusion	158
6.8	Appendix	160
	Bibliography	164

List of Figures

2.1	Treatment	7
2.2	Oil rich provinces of Kazakhstan	18
2.3	Trend	20
2.4	Dependent variable: Unemployment (full year)	23
2.5	Dependent variable: Yearly real logged household income	23
2.6	Share of National Oil Production	29
2.7	Share of National Natural Gas Production	29
3.1	Quarterly Oil Price in US\$.	42
3.2	Income Satisfaction	45
3.3	Oil rich provinces of Kazakhstan	48
3.4	Households' satisfaction with monetary income within the last three months (reported by household head).	55
3.5	Total real household income in Tenge	55
3.6	Coefficients β_k of lag k from estimating specification 3.1	57
4.1	Quarterly Oil Price in US\$ per Barrel (Cushing, OK WTI Spot Price FOB)	77
4.2	Oil rich provinces of Kazakhstan	85
4.3	Proportion of HH reporting the payment of tuition fees.	90
4.4	Employment in the Formal Sector	90
4.5	Real Quarterly Wage in the Formal Sector (logged)	92
4.6	Probability that a household pays tuition fees.	93
4.7	Engel Curve in 2001	96
4.8	Engel Curve after 2001	97

4.9	Share of National Oil Production	105
4.10	Growth rates of the number of students enrolled in technikums in Kazakhstan, based on official statistics	109
4.11	Growth rates of the number of students enrolled in universities or equivalent educational institutions in Kazakhstan, based on official statistics	110
5.1	Output, Employment and Productivity	114
5.2	Treatment and Control Group	125
5.3	Mines and Miners	125
5.4	Persistence	132
5.5	Location of Active Mines	135
5.6	Spatial spillover	136
6.1	Upstream Sector in the Oil and Gas Industry	146
6.2	Number of wells drilled by company	150
6.3	Growth of the constructed time series	151
6.4	Responses to One-Standard-Deviation Structural Shocks	155
6.5	Responses to One-Standard-Deviation Structural Oil Price Shock: Top row rep- resents the response in lower income countries (GDP per capita < 3000). Bot- tom row represents the response in upper income countries (GDP per capita > 12000)	157
6.6	Number of wells drilled	160
6.7	PDF of variable costs (Million \$US, logged)	162
6.8	PDF of total depth (metres, logged)	162
6.9	Number of wells drilled by level of development	163
6.10	Costs of wells drilled by level of development	163

List of Tables

2.1	Share of Employment in different sectors an average between 2002 and 2005.	10
2.2	Unweighted means, standard deviations and weighted means of the variables on the household level.	30
2.3	Unweighted means, standard deviations and weighted means of the variables on the individual level.	31
2.4	Accumulative Production of main oil and gas consortia between 2001-2005 (Oil and Gas in 1000 tonnes), ownership and location.	32
2.5	Accumulative Production of main oil and gas consortia between 2001-2005 (Oil and Gas in 1000 tonnes), ownership and location. (Cont. Table 4)	33
2.6	Employment by major oil and gas producers	33
2.7	Dependent variable: Unemployment (Binary Variable)	34
2.8	Dependent variable: Yearly logged real wage	34
2.9	Dependent variable: Yearly real logged household income	34
2.10	Dependent variable: Minimum income necessary not to feel poor.	35
2.11	Dependent variable: Yearly real logged household Expenditure	35
2.12	Dependent variable: Yearly real logged household income	35
2.13	Dependent variable: Yearly real logged household Expenditure	36
2.14	Dynamics	36
2.15	Dependent variable: Receipt of social benefits (Binary Variable)	36
2.16	Reduced sample (only individuals/households which were part of the survey since 2001)	37
3.1	Conflicts across major oil and gas producers	43

3.2	Difference-in-Differences $(\bar{x}_{2005,ORP} - \bar{x}_{2001/2,ORP}) - (\bar{x}_{2005,OPP} - \bar{x}_{2001/2,OPP})$ for self-reported health condition and the level of education.	59
3.3	Difference-in-Differences $(\bar{x}_{2005,ORD} - \bar{x}_{2001,ORD}) - (\bar{x}_{2005,OPD} - \bar{x}_{2001,OPD})$ for minimum level of income necessary not to feel poor as reported by the house- hold head.	60
3.4	Unweighted means, standard deviations and weighted means of the main vari- ables.	62
3.5	Unweighted and weighted means of the main dummy variables (Cont. Table 3.4).	63
3.6	Accumulative Production of main oil and gas consortia between 2001-2005 (Oil and Gas in 1000 tonnes), ownership and location.	64
3.7	Accumulative Production of main oil and gas consortia between 2001-2005 (Oil and Gas in 1000 tonnes), ownership and location. (Cont. Table 3.6) . . .	65
3.8	Dependent variable: Satisfaction with Household Income	66
3.9	Dependent variable: Satisfaction with Household Income (weighted)	67
3.10	Dependent variable: Satisfaction with Household Income (reduced sample) . .	68
3.11	Dependent variable: Satisfaction with Household Income	69
3.12	Dependent variable: Growth in Satisfaction with Household Income (depen- dent variable in First Differences)	70
3.13	Dependent variable: Logged Real Household Income	71
4.1	Shares of total household expenditure on different types of goods across dis- tricts and years.	89
4.2	Dependent variable: Binary Variable indicating expenditure on tuition fees for tertiary education	105
4.3	Accumulative Production of main oil and gas consortia between 2001-2005 (Oil and Gas in 1000 tonnes), ownership and location.	106
4.4	Accumulative Production of main oil and gas consortia between 2001-2005 (Oil and Gas in 1000 tonnes), ownership and location. (Cont. Table 4.3) . . .	107

4.5	Description, mean and standard deviation of the main variables on the individual and the household level.	108
4.6	Number of Technikums/ Universities in oil-rich provinces of Kazakhstan	109
5.1	Dependent variable: Employment in primary sector	127
5.2	Dependent variable: Population and Employment	127
5.3	Dependent variable: Employment across Industries	128
5.4	Dependent variable: Employment of Males and Females	131
5.5	Dependent variable: Employment of Males and Females across Industries . . .	131
5.6	Descriptives	136
5.7	Dependent variable: Females in Manufacturing	137
5.8	Dependent variable: Permanently Sick and Retired	137
5.9	Dependent variable: Population and Participation	138
5.10	Dependent variable: Employment in Manufacturing	138
6.1	Descriptive statistics of the main variables	152
6.2	Companies in the sample	161

Chapter 1

Introduction

This thesis consists of five stand-alone chapters. Each chapter empirically evaluates questions relating to the life cycle of natural resource extraction. We use three different data sets to shed light on the local impacts of natural resource booms and busts from the local population's perspective (chapter 2-5) and from the firm's perspective (chapter 6).

In chapter 2 to 4, we use the Household Budget Survey of Kazakhstan to evaluate the impact of the recent oil and gas boom on the population located in the proximity of oil and gas fields. To do this, we construct a data sets on the location of the biggest oil and gas fields in Kazakhstan and link it spatially to the Household Budget Survey. Our identification strategy relies on the exogenous (to a small open economy) increase in the price of oil. In chapter 2 we evaluate the local impacts of the oil boom on the income distribution. We find that unemployment and income inequality decreased and average income increased during the boom. In chapter 3, we assess how the increase in average income affected households' satisfaction with income. We show that, even though average income increased, households' satisfaction with income decreased during the boom. We argue that this is driven by inflated expectations regarding the positive impact of the oil and gas boom on households' economic conditions. In chapter 4, we evaluate the impact of the oil boom on household expenditure. We find that the proportion of households reporting the payment of tuition fees for tertiary education increased during the boom. We argue that this is the result of the specific conditions in Kazakhstan, combining a relatively well educated labour force, challenging conditions of

oil and gas extraction which require a highly educated labour force, and a link between the two. In such an environment an oil and gas boom naturally triggers an increase in incentives to invest in tertiary education.

In chapter 5, we study the long term impacts of the coal bust in the UK since 1981. We construct a new data set on the location of coal mines and spatially link it to the UK Census on the district level. The dramatic run down of the coal industry provides us with the required variation to evaluate the long term impacts of a negative labour demand shock on local labour markets. We show that districts affected by coal-mine closures experienced a permanent decrease in population and labour force participation. Interestingly, despite predominantly males being laid-off, we find that female labour force participation decreased. Moreover, we show that female employment in manufacturing persistently decreased in affected districts, whereas male employment in manufacturing increased. We interpret that as evidence for a substitution of females in manufacturing.

In chapter 6, we study the determinants of drilling costs and their impact on the real price of oil. We use a new global data set from Wood Mackenzie providing micro data on the number of exploration wells drilled and unit costs of drilling. We use the data in combination with the real price of crude oil to estimate a three dimensional structural VAR. The proposed structural model of the upstream sector allows us to decompose the variation in the reduced form errors of the estimated VAR into three structural shocks and estimate the dynamic effects of these shocks on the variables in the system. We show that the upstream sector in the oil and gas industry is subject to increasing costs. We also find that shocks to costs of drilling do not affect the real price of oil permanently.

Chapter 2

Distributional Effects of Natural Resource Booms: Evidence from Kazakhstan

2.1 Introduction

Surprisingly little is known about the distributional effects of natural resource booms. This is mainly due to a data availability constraint on income distributions for resource rich countries (Ross 2007; Van der Ploeg 2011). The aim of this paper is to contribute to filling this gap.

To our knowledge, this paper is the first to use data from a household budget survey in a developing country to provide evidence on the local distributional effects of an oil boom. To do this we exploit the exogenous increase in the oil price to a small open economy which affected households in the oil rich region and the oil poor region differently. To evaluate the impacts on the local population we use the Household Budget Survey of Kazakhstan (2001- 2005). The survey has been conducted on a quarterly basis by the Statistical Agency of the Republic of Kazakhstan (SARK). The questionnaire contains detailed information on total household income, household expenditure, employment and education.

Our results suggest that the oil boom in the oil rich districts of Kazakhstan between 2001

and 2005 had four principal effects. First, we find that unemployment decreased. Second, we find that average household income increased and income inequality decreased. We argue that the decrease in income inequality has been driven by the increase in labour demand which in particular benefited households located at the lower end of the income distribution. Finally, the expected positive effect on wages and local inflation remained absent because of the abundant supply of input factors in the oil rich districts of Kazakhstan.

Our results highlight the importance of linkages to the regional labour market. We expect more developing countries to join the group of resource extracting economies (Ross 2012). These discoveries will affect the level of development in those countries positively if the resource extracting sector is sufficiently strongly linked to the regional economy.

This paper builds on a currently active empirical literature (Black et al. 2005a; Marchand 2011; Goderis and Malone 2011; Caselli and Michaels 2013; Aragón and Rud 2013; Allcott and Keniston 2013) which evaluates a well-established theoretical literature (Hirschman 1958; Cordon and Neary 1982; Auty 2001) on the links between resource extractive industries and the local labour market.

Hirschman (1958) emphasizes the importance of forward and backward linkages for economic development. He points out that in particular "primary products from mines, wells and plantations may slip out of the country without leaving much of a trace in the rest of the economy". This might occur if firms operate as enclaves in a developing economy because scope to use local inputs is often restricted by the level of human capital. Moreover, large amounts of revenues leak abroad to service FDI. This reduces local demand and thus the potential for economic development (Auty 2008). The host country could introduce Local Content Requirements¹ (LCRs) to encourage the use of local inputs and reduce the outflow of capital. If implemented, LCRs may induce the expected positive effect on household income and economic development (Aragón and Rud 2013).

¹LCR basically requires companies to use local content of work, goods and services in their production process (Kaiser and Pulsipher 2007)

Cordon and Neary (1982) suggest a way to think about energy booms and labour markets in small open economies with full employment. They differentiate between two principal effects. First, a resource movement effect referring to the fact that during a boom labour moves out of the manufacturing into the energy sector. Second, the spending effect referring to the idea that money entering the economy increases absorption and, thus, production and prices of non-tradable goods. Both effects pull workers out of the tradable goods sector resulting in de-industrialization (Dutch Disease). In a world with perfect labour mobility across sectors, limited migration across countries and full employment this model predicts an increase in average wages and thus household income.

Several papers have recently evaluated these predictions empirically. Marchand (2011) and Black et al. (2005a) explore the effects on regional labour markets caused by an energy price boom. Marchand (2011) identifies a positive labour demand shock, which increased employment opportunities and earnings in the energy extracting industry and resulted in positive spillovers into non-energy sectors. Consistent with Cordon and Neary (1982), Black et al. (2005a) find evidence of modest employment spillovers into sectors with locally traded goods, but not into nationally traded goods. Allcott and Keniston (2013) find only weak evidence for the Dutch Disease and instead argue that oil and gas booms and employment in manufacturing are pro-cyclical. They argue that this is due to the increase in demand for locally traded manufacturing goods, and the creation of linkages between the oil and gas sector and the manufacturing sector. Aragón and Rud (2013) find that, due to the created backward linkages of a large Peruvian goldmine, a 10% increase in the mine's purchases increased average household income by 1.7%. Goderis and Malone (2011) present some cross-country evidence on resource booms and inequalities. They find that following an oil boom inequality decreases in the short run but then returns to its long-run value as the economy grows.

If the government is able to capture a share of the increased revenues, the effect on the regional household income distribution additionally depends on the willingness of the government to invest and redistribute windfalls. The government could use the windfalls to increase public employment, which could result in an increase in average household income and a

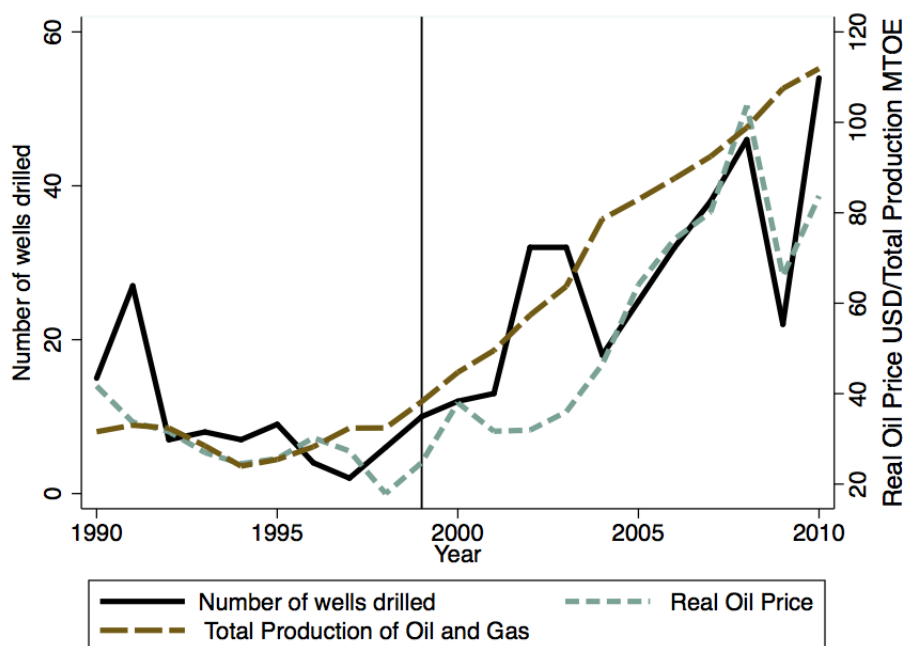
decrease in income inequality. Milanovic (2001) provides evidence that a larger fraction of public employment is negatively associated with income inequality as governments compress income inequality across their employees. Obviously, an increase in social benefits increases average household income and decreases inequality by supporting those located at the lower end of the income distribution. On the other hand, average household income and inequality may increase if oil revenues disappear in the household income of officials as is suggested by Caselli and Michaels (2013).

This paper is structured as follows. The second section provides background information on the oil boom and the creation of linkages to the local economy. Testable propositions are derived from a theoretical framework in section three. The fourth section provides a description of the data employed in the empirical analysis. Our identification and estimation strategy are discussed in section five. The results of the empirical analysis are presented in section six. In section seven we discuss the impact of alternative hypotheses before we conclude.

2.2 Oil Boom and Linkages to the Local Economy

Since the dissolution of the Soviet Union in 1991 until 1999 Kazakhstan's macroeconomic development has been rather poor. This poor economic performance is usually attributed to difficulties to transit from a planned to a market economy, the Russian crisis in 1998, and the exogenous development of commodity prices (Pomfret 2005). Most importantly, the low price of oil did not allow the country to benefit from resource windfalls. As shown in Figure 2.1, the situation changed at the end of 1999. Oil prices started to increase and had roughly quintupled by 2008. Following the increase in the oil price, drilling activity increased significantly. In particular, according to Wood Mackenzie the total number of exploration and appraisal wells drilled in Kazakhstan more than doubled between 2001 and 2002 from 13 to 32. Additionally, the volume of oil and gas produced increased from 40 MTOE (Million Tones of Oil Equivalent) in 1999 up to more than 80 MTOE in 2005.

Figure 2.1: Treatment



Between 1999 and 2010 GDP per capita in Kazakhstan increased from 1130 US\$ in 1999 up to over 9136 US\$ (SARK 2011). Experiencing a nearly tenfold increase in GDP per capita, Kazakhstan joined the group of upper middle income countries. This increase in GDP per capita has been mainly driven by the oil and gas rich Oblasts² (henceforth oil rich provinces or ORP). Only ORPs experienced a more than tenfold increase in GDP per capita.³ Between 1999 and 2009 the share of the oil and gas sector in total GDP increased from 10.9% up to 22.3%. The share of oil and gas in total exports of goods and services increased from slightly above 50% in 2000 up to 58% in 2010 (IMF 2005; IMF 2009).

Huge net revenues have been generated by the oil sector in Kazakhstan. Between 2000 and 2008 net export revenues increased from 2 billion US\$ up to 35 billion US\$. These revenues flowed mainly (1) abroad to serve FDI, (2) to the central and regional governments; (3) to those directly employed in the oil and gas sector; (4) to local firms through the creation of forward and backward linkages. A discussion on (1) and (2) is provided by Esanov and Kuraltbayeva (2008), Najman et al. (2005) and is not covered in this paper. The focus here is

²First level of local government.

³Except Astana and Almaty the new and old capitals, respectively.

on (3) and (4), with a brief discussion of local public expenditures on social benefits.

The quick recovery of the oil and gas industry required an inflow of FDI. The annual volume of FDI has been increasing since 1998 with 1.2 billion US\$ up to 18.4 billion US\$ in 2009. The share of FDI going into the oil and gas industry was initially as high as 74%. In recent years this share has stagnated around 20% (SARK 2011). Consequently, the ownership structure of consortia⁴ in Kazakhstan's oil and gas sector has changed continuously since independence (Kaiser and Pulsipher 2007). In 2004, eleven consortia accounted for nearly 95% of oil and gas production in Kazakhstan (see Tables 2.4 and 2.5 in the Appendix). It is apparent from these Tables that activity in the oil and gas sector has been dominated by International Oil Companies (IOC). Most of the major IOC have been involved in the extraction of oil and gas in Kazakhstan including BG Group, BP, Chevron, ConocoPhillips, ENI-Agip, ExxonMobile, RoyalDutch Shell and Total. Nevertheless, the Kazakh government remains an important player in the oil and gas sector. The state-owned company Kazmunaygas holds an important share of total oil production (15%). Additionally Kazmunaygas is involved in several consortia, most notably in Tengizchevroil with a share of 20%. Expanding its influence, Kazmunaygas also purchased Mangystaumunaigas and a share of Karazhanbasmunay in 2007 and 2009, respectively. Overall, Tables 2.4 and 2.5 suggest that over 60% of production in 2004 was controlled by IOCs, whereas at least 25% of the oil and gas sector was kept under the control of the government.⁵

The dominant role of IOCs in the oil and gas sector is not necessarily a blessing. Since Hirschman (1958), it is well-known that oil and gas companies do not necessarily trigger local economic development but can operate as enclaves without any link to the local economy. Thus, Kazakhstan's government formally and informally tried to enforce LCRs. The 1995 Petroleum Law required subcontractors to be Kazakhstani, and the 1996 Law on Subsurface and Subsurface Use set requirements for using local goods and services (Osanova 2012). Additionally, the government has publicly urged IOCs to comply with LCRs since 2001

⁴Consortium: a partnership of more than one individual, firm or government (or any combination of these) with the mission to achieve a common goal by participating in a common activity and pooling resources.

⁵Central Asia Petroleum Ltd. has an unknown ownership structure and is believed to be strongly tied to President Nazarbaiyev's family.

(Panorama 2012).

Table 2.1 suggests that IOCs directly employ from the local labour market. Here we compare the provincial employment shares of Mangystau and Atyrau, the core of the oil rich region (see Figure 2.6 and 2.7 in the Appendix), with two oil poor provinces (OPP).⁶ The two OPPs Karaganda and South Kazakhstan have been chosen as comparisons. Karaganda is a manufacturing dominated province and South Kazakhstan is dominated by the agricultural sector. Table 2.1 suggests that the employment share in the manufacturing sector is significantly above the national average in Mangystau (31%) and slightly above in Atyrau (14%). Additionally, the share of industrial workers active in the extractive manufacturing is indicated in brackets. The latter suggests that approximately every 15th and every 5th individual of the employed labour force is active in the oil sector in Atyrau and Mangystau respectively. Information from Table 2.6 (see Appendix) additionally confirms that total employment of the biggest consortia adds up to approximately 100000 employees or 6% of the active labour force in ORPs.⁷ We argue that the possibility of employing from the local labour market has been driven by the high level of human capital which Kazakhstan has inherited from the Soviet Union (Yessenova 2008).

In terms of indirect employment, Auty (2008) points out that forward and backward linkages have developed to the regional economy. For instance, forward linkages are developing in Atyrau, the capital of Atyrau Province, which has turned into an important petrochemical producer. And backward linkages are developing as IOCs promote economic activity by encouraging small- and medium-sized enterprises to supply inputs at an appropriate standard and cost. For instance, IOCs directly invest locally to produce plastic pipes which are used as an input factor of oil and gas extraction. The value of the local production of oil and gas equipment by Kazakh national companies tripled between 2001 and 2005 (Kaiser and Pulpispher 2007). Up to 2005 the share of the total market size in oil and gas equipment remained relatively small at 2-3%. However, in absolute terms, 100 million US\$ may have a significant

⁶The focus on these two provinces is driven by data constraints since aggregated data on the district level was not available.

⁷The active labour force (app. 1.8 Mio) in the ORP is estimated using total population of the ORPs and multiplying it by the share of the population economically active on the provincial level (around 65%) (SARK 2009).

Table 2.1: Share of Employment in different sectors an average between 2002 and 2005.

Provinces	Agriculture and Fishery	Manu- facturing*	Construction	Trade	Transport and Commu- nication	Services
Kazakhstan	34%	12% (20%)	5%	15%	7%	27%
Atyrau	14%	14% (46%)	13%	15%	11%	33%
Mangystau	3%	31% (73%)	10%	12%	12%	33%
Karaganda	24%	27% (10%)	4%	13%	7%	26%
South khstan	Kaza- 45%	6% (13%)	4%	17%	6%	22%

Notes: *Resource Extraction as % of employment in the manufacturing sector. Source: SARK (2009).

effect in creating employment opportunities in the sparsely populated region.⁸ Table 2.1 provides additional support for the hypothesis that indirect employment played an important role. The construction sector in the oil rich provinces is large relative to other provinces in Kazakhstan. On average, 52% and 67% in Atyrau and Mangystau, respectively, of the provincial investment in fixed capital between 2001 and 2004 originated from the manufacturing sector which itself is dominated by the oil and gas sector (SARK 2005). Detailed data on activities in the Transportation and Communication sector was not available. But oil and gas are almost exclusively the only tradable goods within the oil rich provinces. Thus, the provincial transportation sector is mainly driven by the requirements of the oil sector to export.⁹ Finally, the service sector employs a relatively large share of those employed in comparison to OPP.¹⁰ As shown by Cordon and Neary (1982), this may be driven by the consequences of the spending effect. Summarizing the existing linkages of major oil and gas projects to the regional economy, Kalyuzhnova (2008) concludes that these projects reinject life into the oil rich districts.

⁸With an area of close to 1 million sqkm, approximately four times as large as the UK, and a population of less than three million, the population density in ORPs is around 3 people per sqkm.

⁹An important example is the sea port in Mangystau. Being one of the biggest harbours in the Caspian region, the Aktau Sea Port is located in the capital of Mangystau, Aktau. The main purpose of the port is the export of oil and gas products. Most recently, oil transports to Baku increased to supply the Baku-Tbilisi-Ceyhan oil pipeline.

¹⁰The service sector is dominated by public sector services (>70%) (SARK 2008). Political and administrative employees of local public administration represent only a small share of the local labour force (1%) (Meruert 2003).

2.3 Theoretical Framework

Inspired by Cordon and Neary (1982), we present a simple general equilibrium framework which is used to derive four testable hypotheses. Based on the discussion above, we assume the oil rich districts of Kazakhstan to consist of two *formal* sectors: a resource extracting sector and a non-tradable goods sector. The agricultural sector and the import competing sector are ignored due to their marginal contribution to the region.¹¹ We additionally introduce an *informal* sector. High living costs and a poorly developed social welfare system make it impossible to be unemployed in the oil rich districts.¹² Nevertheless, a large proportion of our sample reported to be unemployed at the beginning of the boom. Thus, unemployed will be treated as if they are active in the informal sector. Unemployment is introduced by assuming a minimum nominal wage, W_{min} , which is paid to workers employed in the formal sector.¹³ Thus we have three sectors: an oil extracting sector, a non-tradable goods sector and an informal sector, indexed $j \in \{o, nt, i\}$, respectively.

All sectors produce output X_j have the productivity A_j and use labour L_j .¹⁴ The production functions in the formal sectors are $X_j = A_j f_j(L_j)$ where $f_j(0) = 0$, $f_j'(\cdot) > 0$, and $f_j''(\cdot) < 0$. The production function in the informal sector differs in two ways. First, $A_{nt} \gg A_i$ due to the absence of complementarities from capital in the informal sector. Second, we do not assume diminishing returns such that $f_i(0) = 0$, $f_i'(\cdot) > 0$ and $f_i''(\cdot) = 0$. Goods produced in the formal non-tradable goods sector and the informal sector are consumed locally and are assumed to be perfect substitutes, $X_{nt} + X_i = D_{nt} + D_i$. The reader may think of restaurants, hair saloons or craft businesses which can either operate formally or informally. Oil, X_o , is ex-

¹¹Employment share in the agricultural sector in Kazakhstan is below 10% in Mangystau and below 15% in Atyrau. Moreover, a large share of employment in the agricultural and fishery sector in Atyrau, approximately 30%, is employed in the fishery sector. This is mainly due to the extreme climate conditions and a lack of water resources for irrigation, making it very difficult to grow anything in these provinces.

¹²Transportation costs and energy costs are relatively high in the oil rich districts of Kazakhstan. Food has to be transported to remote areas which increases the price of goods. Energy is necessary throughout a significant part of the year to heat dwellings.

¹³Despite the existence of the minimum wage in Kazakhstan, we prefer to think of the minimum wage as being the minimum socially acceptable wage paid (e.g. necessary to preclude civil commotions).

¹⁴We abstract from capital to focus on the main mechanisms. However, capital can easily be introduced in the formal sector when assuming that capital is sector-specific and Kazakhstan is a small open economy which is subject to an exogenous interest rate and an infinite supply of capital.

ported to the rest of the world and manufacturing goods are imported, D_m . Thus, two relative prices determine the equilibrium. $\psi = P_m/P_{nt}$, the relative price which matters to domestic consumers; and $\xi = P_m/P_o$, the terms of trade.

Agents are homogenous, supply one unit of labour and are perfectly mobile across sectors. Total employment in formal and informal sectors is, for the time being, assumed to be constant and given by $\bar{L} = L_o + L_{nt} + L_i$.¹⁵ Profit maximizing firms in the formal sectors employ labour up the point $A_j f'_j = W_{min}/P_j$. The real wage is defined in terms of imports, $w_{min} = W_{min}/P_m$, and is independent of local economic conditions. The remaining workers are employed in the informal sector and are paid $P_{nt}A_i = W_i < W_{min}$. Thus, we have two different wages: a wage in the formal sector and a wage in the informal sector which are indexed $k \in \{min, i\}$, respectively. This allows for an analysis of inequality in real wages and household income. It is important to note that we are mainly interested in the real wage from the wage earners' perspective; a real wage, w_k^* , which takes into account prices of traded and non-traded goods. In our model agents consume only two types of goods, non-tradable and imported. The price of the latter is fixed, such that changes in w^* will be exclusively driven by the price of non-traded goods.

It is assumed that profits in the formal sector, π_o and π_{nt} , are not spent locally but flow either abroad to serve FDI or to other regions of Kazakhstan. Thus, total real absorption is given by $(L_o + L_{nt})w_{min} + L_i w_i = D_m + \psi^{-1}(D_{nt} + D_i) = A$. To close the model, agents are assumed to have constant returns to scale Cobb Douglas preferences and spend a constant share α_m of absorption on manufacturing goods.

We derive our first hypothesis by totally differentiating the equation determining total labour supply $L_i = \bar{L} - L_o - L_{nt}$ with respect to ξ :

¹⁵This assumption will be relaxed and discussed in the empirical section.

$$\frac{dL_i}{d\xi} = \overbrace{\frac{\partial L_i}{\partial \bar{L}} \frac{d\bar{L}}{d\xi}}^{\Delta \bar{L} = 0} + \overbrace{\frac{\partial L_i}{\partial L_o} \frac{dL_o}{d\xi}}^{\text{Resource Movement Effect}} + \overbrace{\frac{\partial L_i}{\partial L_{nt}} \frac{dL_{nt}}{d\xi}}^{\text{Spending Effect}} \quad (2.1)$$

Due to our assumption of a fixed labour supply, the first term is zero. Intuitively, relaxing this assumption would imply an increase in labour supply increasing informal employment. This is because most people migrating into the oil rich districts are initially unemployed and in search of a formal occupation. Thus, we expect the first term on the RHS to be negative. This idea is discussed further in the empirical section. The second and third term are known as the resource movement effect and the spending effect, respectively (Cordon and Neary 1982). The resource movement effect implies that an improvement in the terms of trade increases employment in the oil and gas sector and directly reduces the number of agents active in the informal sector. The spending effect implies that an improvement in the terms of trade increases spending on non-tradable goods. This increases employment in the non-tradable goods sector and thus reduces the share of labour active in the informal sector. Consequently, the second and third terms are positive.¹⁶

Hypothesis 1: An improvement in the terms of trade affects unemployment negatively.

To link individual wages to household income, we assume that every household has one income provider.¹⁷ A decrease in unemployment increases total income, $(L_o + L_{nt})w_{min} + L_i w_i$, because $w_{min} > w_i$. If $\Delta \bar{L} = 0$ average household income also increases and income inequality unambiguously decreases. To see this, we normalize \bar{L} to 1, define $\omega = w_i/w_{min}$ and, based on standard calculations, define the Gini-coefficient as $G = L_i - \omega L_i / ((1 - L_i) + \omega L_i)$. Totally

¹⁶ Proof. Totally differentiate $A_o f'_o = w_{min} \xi$ and $A_{nt} f'_{nt} = w_{min} \psi$ to get:

$$\begin{aligned} A_o f''_o dL_o &= dw_{min} \xi + d\xi w_{min} \\ A_{nt} f''_{nt} dL_{nt} &= dw_{min} \psi + d\psi w_{min} \end{aligned} \quad (2.2)$$

By construction $dw_{min} = 0$. Thus, rearranging the first equation gives $dL_o/d\xi = w_{min}/(f''_o A_o)$ and rearranging the second gives $dL_{nt}/d\xi = (d\psi/d\xi)(w_{min}/(A_{nt} f''_{nt}))$. The first expression is negative because of the assumed concavity. As shown further below, $d\psi/d\xi > 0$ such that the second expression is also negative.

¹⁷ Mode and median of income providers per household in our sample is between one and two.

differentiating the latter with respect to ξ :

$$\frac{dG}{d\xi} = \frac{\partial G}{\partial L_i} \frac{dL_i}{d\xi} + \frac{\partial G}{\partial \omega} \frac{d\omega}{d\xi} > 0 \quad (2.3)$$

We have shown above that $dL_i/d\xi > 0$. $d\omega/d\xi < 0$ because w_{min} is fixed whereas w_i increases because ψ decreases. Thus, both terms are positive such that income inequality decreases.¹⁸

Hypothesis 2: An improvement in the terms of trade affects average household income positively and income inequality negatively.

An important mechanism of the Dutch Disease often accompanying resource booms is the appreciation of the real exchange rate. This is driven by the requirement of the non-traded goods sector to expand production, which in turn is triggered by the spending effect. This may be seen by deriving an expression for the supply of non-tradable goods and totally differentiating in respect to ξ :

$$\frac{d\psi}{d\xi} = \frac{\partial \psi}{\partial X_{nt}} \frac{dX_{nt}}{d\xi} + \frac{\partial \psi}{\partial w_{min}} \frac{dw_{min}}{d\xi} > 0 \quad (2.5)$$

$dX_{nt}/d\xi < 0$ and $dw_{min}/d\xi = 0$ by construction: the former because we assumed that a constant share of absorption $(1 - \alpha_m)$ is spent on non-tradable goods; the latter because we assumed the wage to be fixed. Thus, the first expression on the RHS is positive because $\partial \psi / \partial X_{nt} < 0$ due to the assumed concavity. The second term on the RHS is zero, indicating that the usual effect on local inflation is diluted due to an elastic supply of inputs.

Hypothesis 3: In the presence of excess supply of inputs, an improvement in the terms of trade

¹⁸Partial derivatives are given by

$$\begin{aligned} \frac{\partial G}{\partial L_i} &= 1 - \frac{\omega}{(1 - L_i + \omega L_i)^2} \\ \frac{\partial G}{\partial \omega} &= -\frac{(1 - L_i)L_i}{(1 - L_i + \omega L_i)^2} \end{aligned} \quad (2.4)$$

has a negligible effect on local prices.

Implicitly we assume that households spend their complete household income in every period, $(L_o + L_{nt})w_{min} + L_i w_i = D_m + \psi^{-1}(D_{nt} + D_i)$. Deaton (1992) suggests that individuals consume cash on hand as long as it is less than expected life time income. Kazakhstan is in an economic transition to a market economy and is experiencing a resource boom. Therefore, it appears reasonable to assume that agents expect an increase in household income and, thus, are consuming their total household income. If we are additionally willing to assume that informal sources of income are *not* reported (or partially reported), we expect to observe an increase in reported household income as agents move into formal employment which will not be matched by an equivalently high increase in reported household expenditure. To see this, we fix ψ and w_i , for the time being, to focus on the increase in formal employment, add time subscripts to the equation determining absorption $((L_{o,t} + L_{nt,t})w_{min} + L_{i,t} w_i = D_{m,t} + \psi^{-1}(D_{nt,t} + D_{i,t}))$ and first difference the whole expression:

$$(\Delta L_o + \Delta L_{nt})w_{min} > \Delta D_m + \psi^{-1}(\Delta D_{nt} + \Delta D_i) \quad (2.6)$$

The reader may think of t as being a period during the oil boom (e.g. 2005) and of $t - 1$ being a period at the beginning of the oil boom (e.g. 2001). The inequality results from our assumption that informal sources of income are not reported such that $\Delta L_i w_i$ is missing. $\Delta L_i w_i$ is negative because agents move out of the informal sector during the boom such that $L_{i,t} w_i < L_{i,t-1} w_i$. Note that the limited increase in expenditure will additionally dampen the spending effect and hence the positive effect on prices of non-tradable goods (Hypothesis 3).

Hypothesis 4: In the presence of informal employment, an improvement in the terms of trade has a greater effect on reported household income relative to the effect on reported household expenditure.

2.4 Data

The data employed in the analysis is from the Household Budget Survey of Kazakhstan (2001-2005). The survey was created by the SARK in cooperation with the World Bank (World Bank 2004). Since then it has been conducted on a quarterly basis. IMF (2003) positively evaluated the survey in 2003. The questionnaire contains detailed information on total household income and expenditure, and employment and wages of all household members.

The data is a stratified multi stage cluster sample consisting of 12000 households in each round, representing 0.3 percent of the total number of households (IMF 2003; World Bank 2004). Our focus on the oil rich provinces of Kazakhstan reduces our sample to 1800 households. The household sample has been selected from a household register, which is based on the 1999 population census. The survey is considered to be representative on both, provincial and national levels. In the first stage, within each province (except Almaty and Astana), areas have been appointed into 4 strata: large cities, medium, small towns, and rural settlements. In the second stage, primary sampling units with at least 150 households have been chosen in each strata. Within each primary sampling unit, households were sampled with sampling probability proportional to the size of the household, and thirty households were listed (10 additional households were listed as replacements).

Officially the data set is categorized as a repeated cross-section but it is a perfectly balanced panel within one year. Approximately 25% of the households are replaced by new households annually. Thus, some households are surveyed only one year, whereas others remained in the survey for more than 4 years. In what follows, quarterly reported expenditure and income, wages and employment are aggregated to the yearly level as we are interested in the medium-term impacts of the boom rather than short-run fluctuations.

The panel structure is subject to attrition. Households and individuals not wishing to participate in the survey for a complete year were dropped. The complete-case method has been chosen because dropping households and individuals resulted in a loss of 3% of the total sam-

ple on the household level and 5% on the individual level. Both numbers represent a small share of the sample and are thus very unlikely to affect the consistency of our estimates.¹⁹

Given the homogeneity in weights, the sample can be considered to be self-weighted. Also, the use of weights is essential when calculating population statistics; the use of weights is ambiguous in the case of establishing causal relationships (Solon et al. 2013). In our case, using weights barely affects the consistent estimation of the coefficients and does not provide any efficiency gains. Thus, we prefer to use the unweighted specification. Weighted results are available upon request.

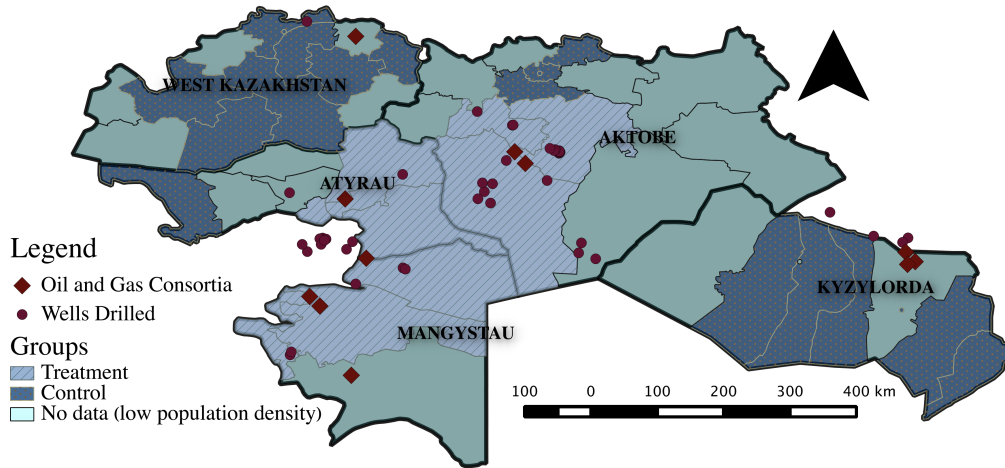
In Tables 2.2 and 2.3 in the Appendix, a description of the main variables may be found. Means and standard deviations are presented for the oil rich districts (ORD) and the oil poor districts (OPD) (see next section for a definition of the oil rich districts). The data on regional inflation has been taken from the SARK (2011), which uses the Household Budget Survey to calculate provincial and national yearly inflation (IMF 2003). The geometric mean of yearly inflation is used to receive real values. The exchange rate at the end of the year 2000 was 144.5 Tenge/US\$, such that the average yearly real household income in the sample is slightly below 2500 US\$ in the oil rich districts and slightly above 1500US\$ in the oil poor districts.

For spatial identification of the resource rich districts, we construct a data set with the location of oil and gas fields which are operated by the biggest consortia (producing in total c. 95% of oil and gas in 2004) in Kazakhstan. We spatially link the location of the oil fields to the location of households on the district level to identify our treatment group. Unfortunately, the SARK was not willing to share the precise location of households. Information on the consortia, most importantly their location, is provided in Tables 2.4 and 2.5 (see Appendix) and Figure 2.2.

¹⁹Additionally, some information is missing on the reported level of household income which is necessary not to feel poor. To resolve this problem, 120 values have been imputed by using the arithmetic mean of the available observations within one year and 4 observations were dropped. The results are robust to dropping the imputed observations.

2.5 Identification and Estimation Strategy

Figure 2.2: Oil rich provinces of Kazakhstan



Notes: The locations of the major consortia listed in Tables 2.4 and 2.5 (see Appendix) have been used to identify treatment and control groups. Consortium: a partnership of more than one individual, firm or government (or any combination of these) with the mission to achieve a common goal by participating in a common activity and pooling resources.

In the empirical analysis we focus on the five ORPs which account for 100% of the oil and gas production in Kazakhstan (see Figures 2.6 and 2.7 in the Appendix). We exclude the remaining provinces from the empirical analysis due to differences in climate, population density and economic structure to preserve homogeneity between the treatment and control groups.²⁰ As a result, our sample consists of 22 Rayons²¹ (henceforth district) and 6 cities²².

The identification strategy employed is essentially a difference in differences. To identify the effects of the oil boom on households located in the oil rich districts of Kazakhstan, we exploit two sources of variation; firstly, the distribution of oil fields across districts. We refer to 3 of the cities and 8 of the districts as the oil rich districts (ORD) or the treated group. The control group or the oil poor districts (OPD) consist of untreated households living outside the oil rich districts (see Figure 2.2). Districts are defined as being oil rich if an oil or gas field

²⁰One could argue that fixed effects could be used to account for these differences. However, fixed effects do not account for different trends which might result from differences in initial conditions and thus violate our identification strategy. We have repeated our analysis using the rest of Kazakhstan as a control group and our results do not differ significantly.

²¹Second level of local government.

²²As defined by the Statistical Agency of Kazakhstan.

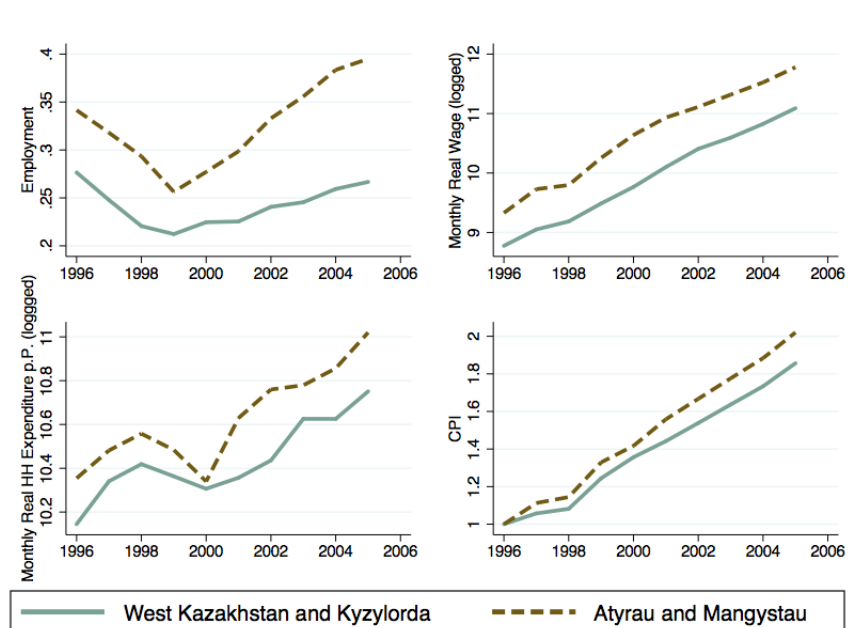
which is operated by one of the the biggest consortia (producing in total c. 95% of oil and gas in 2004) is located within a district. A disadvantage of our identification strategy is that we cannot assign the location of oil fields to individual cities. This is mainly because oil and gas fields are typically not located close to cities. But cities significantly benefit from the oil boom due to the creation of linkages (backward and forward) and the spending effect. Thus, we add Atyrau and Aktau, the two cities, which are located in the core of the oil rich region and are surrounded by oil rich districts, to our treatment group. There is plenty of evidence that both cities significantly benefit from the oil boom (Auty 2008; Pomfret 2005). Dropping the cities does not significantly affect our results.

Secondly, our treatment is represented by the variation in oil price, drilling activity and production over time (see Figure 2.1). Changes in individual measures have different causes. An increase in the global oil price increases drilling activity in the oil and gas sector (Toews and Naumov 2014). This is particularly true for Kazakhstan, where a profitable extraction of oil and gas requires a relatively high price because extraction and transportation costs are relatively high.²³ The increased drilling activity is obviously associated with an increased discovery of oil and gas reserves - most notably Kashagan, one of the biggest global oil discoveries in recent years. We consider the increase in the oil price and the subsequent increase in drilling activity as exogenous to economic outcomes at the household level in a small open economy.

Additionally, production has been increasing since 1999. It is more difficult to argue that the increase in production is exogenous to economic outcomes at the household level. This is because standard economic theory suggests that output and wages are determined by workers' productivity. We argue that the increase in production has been mainly driven by an increased inflow in FDI which allowed for a continuous improvement of production facilities

²³First, the climate conditions which make extraction of oil and gas extremely difficult and thus very costly. Kashagan is now considered to be the most expensive project in the world (Hargreaves 2013). Second, Kazakhstan is landlocked such that pipelines have to be built to transport oil and gas to other markets. For instance, PetroKazakhstan is located in Kyzylorda - a land locked region within a land locked country - where transportation costs have been estimated to lie around 12 US\$ per barrel (Pomfret 2006). 12 US\$ corresponded approximately to the price of a barrel at the beginning of 1999.

Figure 2.3: Trend



and a relaxation of constraints to export oil and gas (Pomfret 2006).²⁴ Unfortunately, we cannot disentangle the impact of the improvement in the terms of trade from the increase in oil and gas production. However, if we interpret the increased production as an exogenous increase in productivity (in the framework of our model an increase in A_o), our theoretical predictions will remain qualitatively identical to the improvement in the terms of trade.

District fixed effects are employed to account for constant unobservables. Year fixed effects are included to account for shocks affecting oil rich and oil poor districts simultaneously. The implicit identification strategy in the difference-in-differences framework is that, in the absence of an oil boom, the development of the outcome variable would be similar in oil rich and oil poor districts and the coefficient of the treatment variable would remain insignificant. Hence, the similarity of trends in both areas in the absence of the oil boom is a necessary condition for the validity of the identification strategy. This strategy would be invalidated in the case of other unobserved time-varying factors correlated with the oil price and differently

²⁴In particular, the construction of a pipeline to connect the biggest onshore oilfield in Kazakhstan, Tengiz (see Table 2.4 in Appendix), to the Black Sea had a significant effect on the production possibilities in the region. In October 2001 Caspian Pipeline Consortium (CPC) loaded crude oil onto a tanker in the Black Sea near the Russian city of Novorossiysk for the first time (CPC 2011).

affecting the oil rich and the oil poor districts. Thus, it is standard practice to look at trends in economic outcomes before the beginning of the treatment. Unfortunately, our sample starts in 2001 which is approximately the beginning of the treatment. So we use aggregated data on the province level from SARK to compare trends between the core of the oil rich provinces (Mangystau and Atyrau) and its periphery (Kyzylorda and West Kazakhstan) for four outcome variables (see Figure 2.3): employment, real wages, real household expenditure and the consumer price index. Optimally, we would use unemployment instead of employment and additionally construct a time series for household income. Unfortunately, the definition of these variables has been subject to changes such that we have not been able to construct a homogenous time series for this time period. Graphs in Figure 2.3 suggest a common trend in all variables before the beginning of the boom around 2000. We observe an obvious divergence in employment since 2000. We do not observe a difference in trends for wages after 2001. A weak divergence can be observed for real household expenditure and CPI from the beginning of 2001. This is consistent with our results. To account for time varying unobservables on the local level, province specific trends and district specific trends are employed in the subsequent empirical analysis.

To test *hypotheses 1-4* formally we employ the following specification:

$$y_{z,d,t} = \beta \Gamma_t D_d + x'_{z,t} \zeta + \alpha_d + \gamma_t + \epsilon_{z,d,t} \quad (2.7)$$

$y_{z,d,t}$ is the outcome variable of $z \in \{h, i\}$ indicating households or individuals, respectively, in district d in period t . On the individual level, the outcome variable is either a binary variable indicating whether an individual reported being unemployed *or* the yearly real wage (logged) (see 2.3 in the Appendix for a definition of the variables). On the household level, the outcome variable is either yearly real household income (logged) *or* real household expenditure (logged) *or* reported amount of household income necessary not to feel poor (logged) (see Table 2.2 in the Appendix for a definition of the variables). The latter is used as an indicator for an increase in local prices.²⁵ D_d is a dummy variable, which takes the value one if an

²⁵In the presence of non-homothetic preferences we may be worried that agents in ORD and OPD care about different bundles of goods and those different prices. In particular, we would be worried that the increase in

individual is located in an oil rich district and zero otherwise; Γ_t is equal to zero in 2001 and is equal to 1 otherwise; $x_{z,d,t}$ is a vector of individual/household head specific characteristics controlling for sex, age, ethnicity and education; $\epsilon_{z,d,t}$ is the individual/household specific error. Assuming that errors may be correlated over time from the same household / individual and across households / individuals from the same district, we cluster at the district level, the highest possible level of clustering (Angrist and Pischke 2009). In the case of quantile regressions the errors are block bootstrapped at the district level.

To exploit the dynamics of the outcome variables, the following specification is estimated allowing β to vary over time:

$$y_{z,d,t} = \sum_{k=2002}^{2005} \beta_k \Gamma_k D_d + x'_{z,t} \zeta + \alpha_d + \gamma_t + \epsilon_{z,d,t} \quad (2.8)$$

In specification 2.8 Γ_k is equal to 1 if $k = t$. This specification allows for an exploration of the average effect of the oil boom on our outcome variables in the individual years relative to 2001.

2.6 Results

The results from estimating equation 2.7 are presented in Tables 2.7-2.11 (see Appendix). In the first column, the most parsimonious specification is presented. In column 2, individual or household specific and time varying observable characteristics are added. In columns 3 and 4, province specific trends and district specific trends are added respectively. In Tables 2.12 and 2.13 (see Appendix) the results from estimating equation 2.7 are presented for total real household income and household expenditure. Instead of estimating the average treatment effects using OLS, results are presented for the individual quantiles using quantile regressions.

household income in ORD of Kazakhstan increases the share of manufacturing goods in the consumption bundle where the price is not determined locally. Toews and Libman (2014) show that this is unlikely to be a concern and that it is more likely that consumption of non-tradables increases, namely, education and services.

Figure 2.4: Dependent variable: Unemployment (full year)

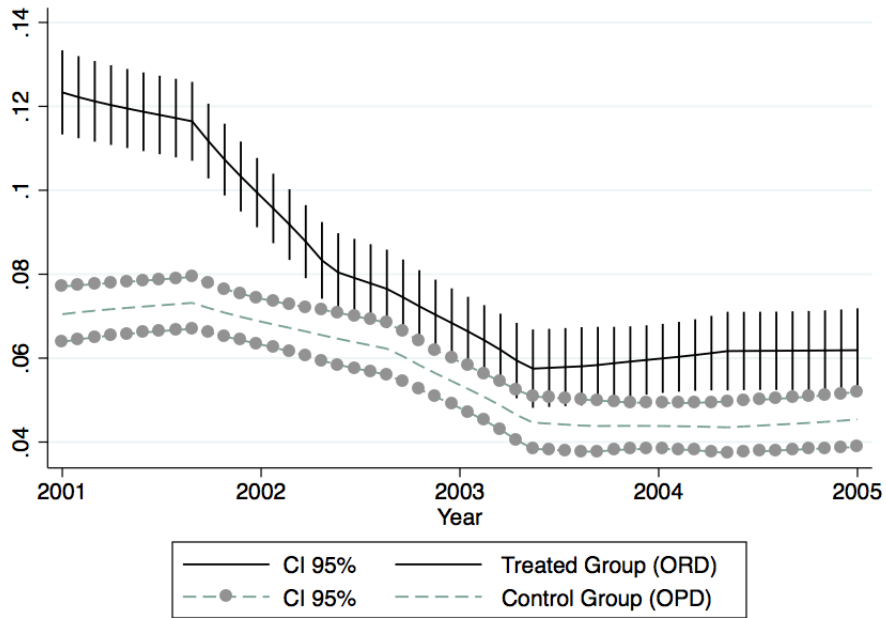
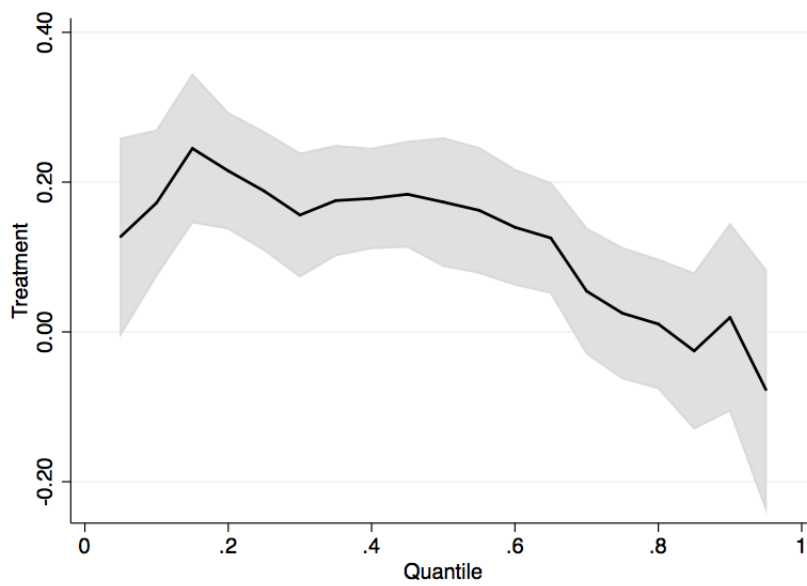


Figure 2.5: Dependent variable: Yearly real logged household income



The results for household income are additionally presented graphically in Figure 2.5. In Table 2.14 the results from estimating equation 2.8 are presented for the variables in Tables 2.7 - 2.11 adding the full set of individual specific controls.

In Table 2.7 the results for the test of the first hypothesis are presented. The results suggest that unemployment had decreased by 5 percentage points since the beginning of the boom. This result is additionally depicted in Figure 2.4. Consistent with our assumption on the existence of a minimum wage, we do not observe a significant effect on the average real wage of those employed in the formal sector (see Table 2.8). However, note that the point estimate is predominately negative, suggesting that the real wage w^* (real wage taking into account prices of tradables and non-tradables) might have slightly decreased. In Tables 2.9 and 2.12 the second hypothesis is tested. The results suggest that since 2001 real household income increased in ORD relative to the OPD by approximately 10%. The effect increases, once province or district specific trends are added to the regression. This might be driven by the fact that we do not control for migration into the oil rich districts. Omitting migration would bias β downwards if mainly poor households migrate into the oil rich districts. Most importantly, disaggregating the average effect into effects on individual quantiles, we find that in particular households located in the lower end of the distribution benefited from the oil boom (see Table 2.12 in the Appendix). Our results suggest that households in the lower end of the distribution experienced an increase in household income by approximately 20%. At the same time, households at the upper end of the distribution did not benefit at all. This result is graphically presented in Figure 2.5. In Table 2.10 we present the result for the test of the third hypothesis. We argued that due to an excess in labour supply and a sufficiently high inflow of FDI the local impact on prices in the oil rich districts of Kazakhstan remained limited. We cannot reject the hypothesis that households in the oil rich districts did not experience a disproportionate increase in local prices. Finally, in Tables 2.11 and 2.13 we present the results of our fourth hypothesis test. If we are willing to assume that households are reluctant to report income from informal sources, our result provides additional evidence that the increase in formal employment is the main channel at work. In particular, even though the point estimates are mainly positive we cannot reject the hypothesis that household expenditure did

not significantly increase in the oil rich districts during the boom. Under the assumption that household income is spent immediately and not saved, this provides additional evidence that households moved into formal employment.

In Table 2.14 (see Appendix) we estimate specification 2.8 which allows us to explore the dynamics of our outcome variables in more detail. The most important insight from Table 2.14 is that the effects on unemployment and household income are stronger in the beginning of the boom. As already argued before this can be driven by the continuous migration into the oil rich districts which should theoretically dilute the negative effect on unemployment and the positive effect on real household income.

2.7 Alternative Explanations and Concerns

Social Benefits

The decrease in household income inequality might have been caused by an increase in social benefits paid to households in ORDs. During the oil boom, regional governments were able to capture a share of the rents (MFRK 2011). If they had been able to redistribute rents to support households at the lower end of the income distribution income inequality might have decreased. To test this hypothesis we reestimate specification 2.7 using a binary variable, which is equal to one if a household reports that social benefits have been received and zero otherwise. The results are presented in Table 2.15 in the Appendix. Consistent with the increase in formal employment, we observe a decrease in the probability of receiving social benefits.

Household Income vs. Household Expenditure

We argued that the reason that we observe a significant effect on real household income, but an insignificant effect on real household expenditure, is due to the fact that households are reluctant to report income from informal sources. To support this hypothesis we conduct a simple t-test and compare average household income and average household expenditure for the individual years (2001-2005). On the 1% level we reject the hypothesis that average

household income and average household expenditure is equal in every year of our sample but for 2004. Interestingly, up to 2004 our results suggest average household expenditure to be significantly greater whereas in 2005 average household income is significantly greater. This is consistent with the idea that households underreported income in times of higher unemployment and thus higher informal activity. Alternatively, households may have either lived off their savings or borrowed money. Neither can be confirmed by the information we have from the household budget survey. Barely anyone reported using savings or borrowing money for current expenditures.

Oil Workers

The survey is conducted with households being residents in an individual house or in part of a house. But foreigners or nationals from other provinces who are employed by the IOCs usually work in shifts. Similar to workers placed offshore on a rig, IOC employees are transported to the oil and gas fields (which are often located in remote areas) where they remain for a certain period before they return home. This results in an under-sampling of those individuals and households who are located abroad or in other provinces since they are not captured by the survey. The under sampling has a serious consequence. The estimated effect of the oil boom on the distribution of income is likely to be biased downwards and all the results should be rather seen as the lower bound.

Migration

An important assumption of our identification strategy is that labour does not migrate. Migration in Kazakhstan is indeed costly due to the size of the country and the difficult climate conditions in the oil rich districts. Nevertheless, we know that oil rich districts have experienced an immigration of workers from other provinces since the beginning of the oil boom. Unfortunately, we are not aware of any data on migration at the district level such that we are unable to control for it. However, there is an alternative to test whether our results are driven by the immigration of households into the oil rich districts. We conduct our analysis by focusing only on households which remained within the sample since the beginning of the

boom in 2001.²⁶ This subsample does not contain households migrating in and out of the oil rich districts after 2000. We reestimate specification 2.7 using the reduced sample and present the results in 2.16. By reducing the sample our estimates of β for the individual specifications does not seem to be affected significantly (even though the effect on unemployment remains statistically insignificant). We interpret that as evidence that our results are not driven by immigration of households from other regions.

2.8 Conclusion

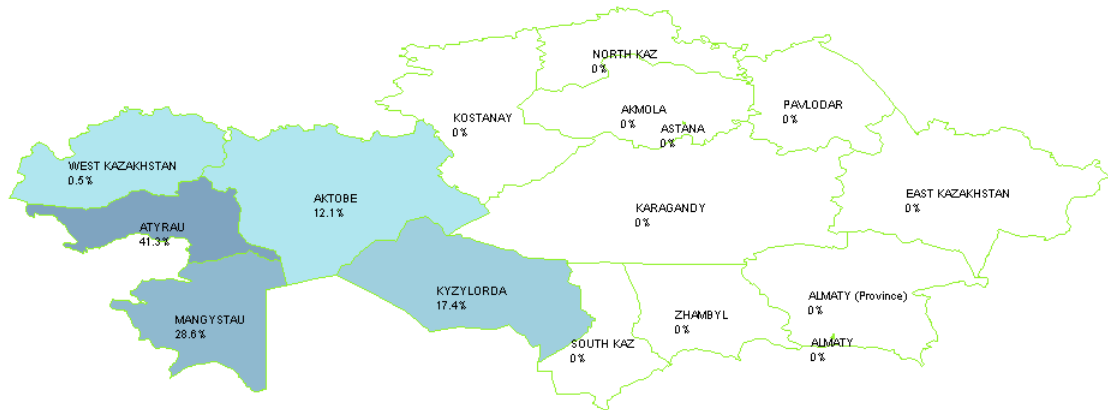
This paper contributes to understanding the effects of an oil boom on the regional household income distribution by providing evidence from the recent oil boom in the oil rich districts of Kazakhstan. To our knowledge, this paper is the first to use a household budget survey to provide evidence of the local distributional effects of an oil boom in a developing country. Our results suggest that, during the oil boom, unemployment decreased significantly by 5 percentage points whereas wages were not affected. As a result of the former, the increase in average household income has been accompanied by a decrease in income inequality. We do not observe a similar effect on average household expenditure. We argue that this is due to the fact that households are always willing to report expenditure but are reluctant to report income from informal sources. Thus, we interpret the significant effect on household income, in combination with a small and insignificant effect on expenditure, as evidence for households moving out of informal into formal employment during the resource boom. Finally, the abundant supply of capital (FDI inflows) and labour (high unemployment) diluted the expected oil boom accompanying increase in local prices.

²⁶Officially, the data set is categorized as a repeated cross section but it is a perfectly balanced panel within one year. Approximately 25% of the households are replaced by new households annually. Thus, some households are surveyed for only one year whereas other households remained in the survey for more than 4 years. The introduced rotation still allows us to construct an unbalanced panel across years by identifying households who remain in the sample. The household ID does not change as long as the household remains part of the survey. Once a household is dropped from the survey the ID is assigned to another household which is newly added to the survey. Thus, we used household characteristics which are unlikely to change to identify the panel dimension of the household. In particular, we used household ID, dwelling type, size of the dwelling, the year the house was built, the sex, date of birth and ethnicity of the household head.

Our results highlight the importance of linkages to the regional economy, as already emphasized by Hirschman (1958). If a sector is linked to the regional economy a boom will increase average household income but also decrease regional income inequality by reducing unemployment or informal employment. The results are very important as we expect more developing countries to join the group of resource-extracting economies (Ross 2012). These discoveries will affect the level of development in those countries positively if resource-extractive sectors are sufficiently strongly linked to the regional economy.

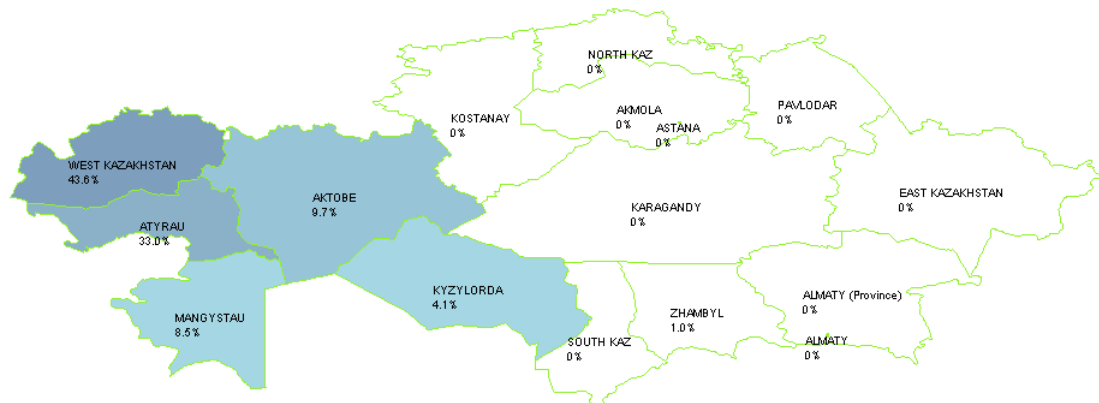
2.9 Appendix

Figure 2.6: Share of National Oil Production



Source:: SARK (2011).

Figure 2.7: Share of National Natural Gas Production



Source: SARK (2011).

Table 2.2: Unweighted means, standard deviations and weighted means of the variables on the household level.

Variables	Description	Mean (ORD ⁺)	SD (ORD)	Weighted Mean (ORD)	Mean (OPD ⁺⁺)	SD (ORD)	Weighted Mean (OPD)
Sample Size	Number of Individual observations Cross-section *Time periods (NT)	3069	-	-	6035	-	-
Real Household Income	Aggregation of yearly household income from various income sources in Tenge. Provincial inflation from the SARK (2011) is used to discount household income to the end of 2000/ beginning of 2001	359070	265332	354853	217752	163481	215747
Real Household Expenditure	Aggregation of yearly household expenditure from various income sources in Tenge. Provincial inflation from the SARK (2011) is used to discount expenditure to the end of 2000/ beginning of 2001.	353649	214024	352548	261775	152257	260938
Minimum income not to feel poor	Households have been asked: Given the current prices, what monthly income should a family have with the same number of people as yours not to feel poor? (Reported in Tenge).	812608	476805	798496	561471	307245	555756
Social benefits	Dummy variable indicating whether a household reported receiving social benefits	0.29	0.45	0.3	0.28	0.45	0.29
Dummy for male	Dummy is equal to one if the household head is a male.	0.61	0.48	0.62	0.56	0.50	0.57
Age	Age of household head.	49	13		50	14	50
High Education	Dummy is equal to one if household head reported to have graduated.	0.13	0.34	0.13	0.2	0.4	0.19
Middle Education	Dummy is equal to one if household head reported to have a vocational education.	0.47	0.5	0.46	0.36	0.48	0.37
Ethnically Kazakh	Dummy equal to one if household head reported to be Kazakh	0.79	0.40	0.82	0.67	0.47	0.66
Dummy for formal employment	Dummy equal to one if household head reported to be employed in the formal sector (public sector employment or employed by a private company) for a complete year	0.54	0.5	0.54	0.43	0.5	0.42

Notes: + oil rich districts, ++oil poor districts.

Table 2.3: Unweighted means, standard deviations and weighted means of the variables on the individual level.

Variables	Description	Mean (ORD ⁺)	SD (ORD)	Weighted Mean (ORD)	Mean (OPD ⁺⁺)	SD (ORD)	Weighted Mean (OPD)
Sample Size	Number of Individual observations Cross-section *Time periods (NT)	8271	-	-	11380	-	-
Unemployment	Dummy equal to one if individual reported to be unemployed for the complete year	0.09	0.28	0.09	0.07	0.23	0.06
Yearly real wage	Individual yearly real wage for the complete year in the formal sector (Individuals who reported being employed in the public sector or working for a company in the private sector). Provincial inflation from the SARK (2011) is used to discount the wage to the end of 2000/ beginning of 2001	210855	164264		128700	82757	
Dummy for male	Dummy is equal to one if the individual is a male.	0.49	0.5	0.49	0.47	0.50	0.47
Age	Age of individual.	34	12	34	36	12	36
High Education	Dummy is equal to one if the individual reported to have graduated.	0.11	0.33	0.11	0.16	0.36	0.15
Middle Education	Dummy is equal to one if the individual reported to have a vocational education.	0.37	0.48	0.36	0.32	0.47	0.32
Ethnically Kazakh	Dummy equal to one if the individual reported to be Kazakh	0.88	0.33	0.89	0.77	0.42	0.76

Notes: + oil rich districts, ++oil poor districts.

Table 2.4: Accumulative Production of main oil and gas consortia between 2001-2005 (Oil and Gas in 1000 tonnes), ownership and location.

Name	Major Owner	Cum. Prod.	Share (in %)	Province	District	Webpage
Kazakhstan	-	265000	100%	Aktobe, Atyrau, Kyzylorda, Mangystau, West-Kazakhstan	-	-
Tengiz-Chevroil	Chevron, Exxon-Mobil, KazMunayGas, LukArco	65500	25%	Atyrau, Mangystau	Zhylyoyskiy (Atyrau), Beyneuskiy (Mangystau)	www.tengizchevroil.com
EmbaMunay Gas	KazMunay-Gas	13000	5%	Atyrau	Kzylkoginskiy, Isatayiskiy, Makhambetskiy, Makatskiy, Zhylyoyskiy, Atyrau (city)	www.kmg.kz
UzenMunay Gas	KazMunay-Gas	27000	10%	Mangystau	Manghystauskiy, Karakhiyanskiy, Zhanozen (city)	www.kmg.kz
Mangystau-MunayGas	Central Asia Petroleum Ltd. (until 2009)	25000	9%	Mangystau	Manghystauskiy, Karakhiyanskiy, Tupkaraganskiy, Zhanozen (city), Aktau (city)	www.mmg.kz
Karazhanbasmunai	Canada's Nations Energy Ltd. (until 2006)	10000	4%	Mangystau	Tupkaraganskiy, Manghystauskiy	www.kbm.kz
Karashaganak	BP Group, ENI	36000	13%	West-Kazakhstan	Burlinski	www.kpo.kz

Source: Munayshy Public Foundation (2005).

Table 2.5: Accumulative Production of main oil and gas consortia between 2001-2005 (Oil and Gas in 1000 tonnes), ownership and location. (Cont. Table 4)

Name	Major Owner	Cum. Prod.	Share (in %)	Province	District	Webpage
Actobe -munaigas	CNPC	24000	9%	Actobe	Temirskiy, galzharskiy, Baiganinskiy	Mu- www.cnpc.ch
Kazakhoil Actobe	KazMunay Gas, LukOil	3500	1%	Actobe	Temirskiy, galzharskiy, Baiganinskiy	Mu- www.koa.kz
Petro - Kazakhstan	Petro - Kazakhstan (until 2005)	23000	9%	Kyzylorda	Syrdarinskiy	www.petro -kazakhstan.kz
Petroleum -Turgai	LukOil, Hurricane Kumkol Munai	13000	5%	Kyzylorda	Syrdarinskiy	www.turgai.kz
KasGer -munai	JSC Yuzh-neftegas, Feba Oil AG, Erbdol Erdgras Gommern	8500	3%	Kyzylorda	Syrdarinskiy	www.kazger -munay.kz

Source: Munayshy Public Foundation (2005).

Table 2.6: Employment by major oil and gas producers

Consortium	Direct Employment	Share of Kazakh employment from total	Source
Tengizchevroil LLP	23255 (including subcontractors) (2005)	ca. 90% (2005)	Yessenova (2008)
Petrokazakhstan	900 (2005)	>90% (2007)	PetroKazakhstan (2012), Najman et al. (2005)
Kazmunaygas JSC	52829 (2007)	Kazakh Company	KazMunaiGas (2012)
Karachaganak Pet. Oper. B.V.	4378 (2008)	>80% (2012)	Karachaganak (2012)
Mangistau-munaygas	4400 (2002)	Kazakh Company	Najman et al. (2005)
CNPC - Actob-munaygas JSC	>21500 (including subcontractors) (2007)	>90% (2007)	CNCP (2012)
Karazhambas-munay	2682 (2004)	Kazakh Company	Karazhanbasmunai (2012)

Table 2.7: Dependent variable: Unemployment (Binary Variable)

	OLS1	OLS2	OLS3	OLS4
	(1)	(2)	(3)	(4)
β	-0.046 (0.022)**	-0.045 (0.022)**	-0.028 (0.021)	-0.046 (0.025)*
Control	N	Y	Y	Y
Province Trend	N	N	Y	N
District Trend	N	N	N	Y
Obs.	21651	21651	21651	21651

Notes: Standard errors adjusted for clustering at the district level are in parentheses : *** p<0.01, ** p<0.05, * p<0.1. District specific effects and time fixed effects are included in every regression. The full set of individual characteristics includes sex, age, ethnicity and the level of education.

Table 2.8: Dependent variable: Yearly logged real wage

	OLS1	OLS2	OLS3	OLS4
	(1)	(2)	(3)	(4)
β	-0.065 (0.056)	-0.088 (0.055)	-0.019 (0.065)	0.015 (0.057)
Control	N	Y	Y	Y
Province Trend	N	N	Y	N
District Trend	N	N	N	Y
Obs.	7683	7683	7683	7683

Notes: Standard errors adjusted for clustering at the district level are in parentheses : *** p<0.01, ** p<0.05, * p<0.1. District specific effects and time fixed effects are included in every regression. The full set of individual characteristics includes sex, age, ethnicity and the level of education.

Table 2.9: Dependent variable: Yearly real logged household income

	OLS1	OLS2	OLS3	OLS4
	(1)	(2)	(3)	(4)
β	0.13 (0.041)***	0.114 (0.041)***	0.174 (0.049)***	0.156 (0.046)***
Control	N	Y	Y	Y
Province Trend	N	N	Y	N
District Trend	N	N	N	Y
Obs.	9104	9104	9104	9104

Notes: Standard errors adjusted for clustering at the district level are in parentheses : *** p<0.01, ** p<0.05, * p<0.1. District specific effects and time fixed effects are included in every regression. The full set of household head characteristics includes sex, age, ethnicity and the level of education of the household head. Household size and household size squared is included in every regression.

Table 2.10: Dependent variable: Minimum income necessary not to feel poor.

	OLS1	OLS2	OLS3	OLS4
	(1)	(2)	(3)	(4)
β	0.01 (0.052)	0.005 (0.051)	0.053 (0.071)	-.014 (0.064)
Control	N	Y	Y	Y
Province Trend	N	N	Y	N
District Trend	N	N	N	Y
Obs.	9104	9104	9104	9104

Notes: Standard errors adjusted for clustering at the district level are in parentheses : *** p<0.01, ** p<0.05, * p<0.1. District specific effects and time fixed effects are included in every regression. The full set of household head characteristics includes sex, age, ethnicity and the level of education of the household head. Household size and household size squared is included in every regression.

Table 2.11: Dependent variable: Yearly real logged household Expenditure

	OLS1	OLS2	OLS3	OLS4
	(1)	(2)	(3)	(4)
β	0.02 (0.06)	0.006 (0.061)	0.046 (0.055)	-.014 (0.063)
Control	N	Y	Y	Y
Province Trend	N	N	Y	N
District Trend	N	N	N	Y
Obs.	9104	9104	9104	9104

Notes: Standard errors adjusted for clustering at the district level are in parentheses : *** p<0.01, ** p<0.05, * p<0.1. District specific effects and time fixed effects are included in every regression. The full set of household head characteristics includes sex, age, ethnicity and the level of education of the household head. Household size and household size squared is included in every regression.

Table 2.12: Dependent variable: Yearly real logged household income

	QR1	QR3	QR5	QR7	QR9
	(1)	(2)	(3)	(4)	(5)
β	0.25 (0.09)***	0.227 (0.07)***	0.142 (0.068)**	0.025 (0.071)	-.088 (0.077)
Control	Y	Y	Y	Y	Y
Obs.	9104	9104	9104	9104	9104

Notes: Standard errors adjusted for clustering at the district level are in parentheses : *** p<0.01, ** p<0.05, * p<0.1. District specific effects and time fixed effects are included in every regression. The full set of household head characteristics includes sex, age, ethnicity and the level of education of the household head. Household size and household size squared is included in every regression.

Table 2.13: Dependent variable: Yearly real logged household Expenditure

	QR1	QR3	QR5	QR7	QR9
	(1)	(2)	(3)	(4)	(5)
β	0.014 (0.106)	0.049 (0.097)	0.069 (0.082)	0.045 (0.083)	-.002 (0.107)
Control	Y	Y	Y	Y	Y
Obs.	9104	9104	9104	9104	9104

Notes: Standard errors adjusted for clustering at the district level are in parentheses : *** p<0.01, ** p<0.05, * p<0.1. District specific effects and time fixed effects are included in every regression. The full set of household head characteristics includes sex, age, ethnicity and the level of education of the household head. Household size and household size squared is included in every regression.

Table 2.14: Dynamics

	Unemployment	ln(wage)	ln(income)	ln(expenditure)	ln(min. income)
	(1)	(2)	(3)	(4)	(5)
β_{2002}	-.038 (0.024)	-.013 (0.05)	0.135 (0.037)***	0.043 (0.052)	-.011 (0.049)
β_{2003}	-.058 (0.023)**	-.088 (0.06)	0.141 (0.044)***	-.072 (0.067)	0.014 (0.066)
β_{2004}	-.041 (0.023)*	-.121 (0.068)*	0.078 (0.056)	0.018 (0.073)	-.006 (0.053)
β_{2005}	-.044 (0.024)*	-.116 (0.064)*	0.102 (0.056)*	0.038 (0.072)	0.022 (0.067)
Control	Y	Y	Y	Y	Y
Obs.	21651	7683	9104	9104	9104

Notes: Standard errors adjusted for clustering at the district level are in parentheses : *** p<0.01, ** p<0.05, * p<0.1. District specific effects and time fixed effects are included in every regression. For column 1-2: : the full set of individual characteristics includes sex, age, ethnicity and the level of education. For column 3-5 the full set of household head characteristics includes sex, age, ethnicity and the level of education of the household head. Household size and household size squared is included in every regression.

Table 2.15: Dependent variable: Receipt of social benefits (Binary Variable)

	OLS1	OLS2	OLS3	OLS4
	(1)	(2)	(3)	(4)
β	-.077 (0.036)**	-.068 (0.035)*	-.027 (0.04)	0.003 (0.037)
Control	N	Y	Y	Y
Province Trend	N	N	Y	N
District Trend	N	N	N	Y
Obs.	9104	9104	9104	9104

Notes: Standard errors adjusted for clustering at the district level are in parentheses : *** p<0.01, ** p<0.05, * p<0.1. District specific effects and time fixed effects are included in every regression. The full set of household head characteristics includes sex, age, ethnicity and the level of education of the household head. Household size and household size squared is included in every regression.

Table 2.16: Reduced sample (only individuals/households which were part of the survey since 2001)

	Unemployment	ln(wage)	ln(income)	ln(expenditure)	ln(min. income)
	(1)	(2)	(3)	(4)	(5)
β	-0.036 (0.028)	-0.100 (0.072)	0.102 (0.049)**	0.004 (0.07)	0.04 (0.053)
Control	Y	Y	Y	Y	Y
Obs.	7778	2618	3482	3482	3482

Notes: Standard errors adjusted for clustering at the district level are in parentheses : *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. District specific effects and time fixed effects are included in every regression. For column 1-2: : the full set of individual characteristics includes sex, age, ethnicity and the level of education. For column 3-5 the full set of household head characteristics includes sex, age, ethnicity and the level of education of the household head. Household size and household size squared is included in every regression.

Chapter 3

Inflated Expectations and Natural Resource Booms: Evidence from Kazakhstan

3.1 Introduction

Maybe the most famous phrase of the US Declaration of Independence ends with "Life, Liberty and the pursuit of Happiness". It requires that everybody has the right to strive for happiness but does not guarantee that everybody can achieve it. The oil boom in the resource rich region of Kazakhstan at the beginning of the 21st century increased local labour demand and average household income.¹ This relaxed constraints to strive for happiness suggesting that locals should since be happier with their monetary income. Unfortunately, this does not appear to be the case as the boom was also accompanied by recurrent local conflicts; most recently the deadly riots in Zhanaozen in December 2011.² We argue that locals remained dissatisfied with the increase in income because it fell short of their expectations (Ross 2007).³

Linking expectations to conflicts, Gurr (1970) suggests that "[m]en are quick to [...] raise

¹See Toews (2014) for a discussion of the distributional effects of the oil and gas boom in Kazakhstan.

²See Table 3.1 for further examples.

³Similar phenomena have already been observed in the past. When documenting the French Revolution, the philosopher Alexis de Tocqueville argued that it had not been triggered by the miserable conditions of the poor but by the swiftly rising expectations of the French people (Tocqueville 1986).

expectations beyond their social means and quick to anger when those means prove inadequate". He concludes that the potential for collective movements varies strongly with unfulfilled aspirations. This result relates to a large literature on natural resource abundance and conflicts (Collier and Hoeffler 2004; Humphreys 2005; Brunnschweiler and Bulte 2008; Frankel 2010). The link between natural resources, aspirations and conflicts suggests that an important channel might be neither greed nor grievances but rather unmet aspirations.

In this paper we explore whether exogenous fluctuations in the oil price negatively affect households' satisfaction with income. We base our analysis on the assumption that income aspirations formed by households located in the oil rich districts of Kazakhstan are a function of the publicly observable oil price, the price of the essentially only locally produced tradable good in the region. Based on this assumption, we argue that the increase in the oil price inflated expectations regarding future income and thus decreased income satisfaction of households located in the oil rich districts of Kazakhstan. To our knowledge, this is the first paper to use a quasi natural experiment to provide evidence on how natural resource booms affect people's satisfaction with income. Our results suggest that the oil boom in the oil rich districts of Kazakhstan between 2001 and 2005 negatively affected households' satisfaction with income. In particular, a 20% increase in the oil prices decreases satisfaction with income by one-tenth of a standard deviation with a lag of half a year. We argue that this is mainly due to inflated expectations.

To study the effect of the oil price boom on satisfaction with income we use the Household Budget Survey of Kazakhstan - a quarterly, unbalanced panel of households covering the period 2001-2005. Of particular interest to the study is that the survey contains information on how satisfied households are with their household income. Satisfaction with household income is reported on the quarterly basis by the household head on a scale from 1-5. For identification we exploit two sources of variation. Spatially, we merge households' locations with locations of major oil and gas fields. Over time we use oil price fluctuations to treat households located in the district of a major oil and gas field, using households located outside these districts as controls.

Our results highlight the importance of managing peoples' expectations in a rapidly changing economic environment. Managing expectations could imply the creation of a narrative which is communicated to the population (Collier 2013). Such a narrative should contain the idea that an increase in commodity prices or the discovery of natural resources will not have any immediate positive effect on the local economic development. Instead, time has to pass for economic adjustments to take place and positive effects to be observed.⁴ The results are very important as we expect more developing countries to join the group of resource extracting economies (Ross 2012). The extraction of natural resources could have a positive impact on the local labour market. However, impacts on the local economy will surely require time. In the meantime, people's expectations will have to be managed in order to avoid increasing dissatisfaction and, more importantly, conflicts.

There appears to be agreement that people are unable or unwilling to make absolute judgments (Frey 2008; Dellavigna 2009). Instead, individuals form aspirations, or equivalently, reference points and use these to assess individual satisfaction with income. In what follows we will build on the happiness literature using aspirations to refer to the formation of reference points - the term famously used in behavioral economics to refer to rationally formed expectations about outcomes (Kahneman and Tversky 1979; Köszegi and Rabin 2006). Thus, satisfaction with income in period t , $s_t = g(y_t, a_t)$, is a positive function of income $\partial g / \partial y_t > 0$ but is *negatively* affected by aspirations $\partial g / \partial a_t < 0$. Since the documentation of the Easterlin Paradox⁵ many authors have focused on the identification of the channels determining aspiration levels. Three channels have been identified as being particularly important in determining the level of aspirations: past income, relative current income and expected future income: $a_t = f(\sum_{k=1}^K y_{t-k} \gamma_k, \sum_{j=1}^J y_j \omega_j, E_{t-x}[y_t])$. The first term is some weighted average of income y over the last K periods. This term captures the idea that an initially positive effect of an increase in income over the last K periods on satisfaction with income vanishes in the long run due to adjusting aspirations (Easterlin 2001). The second term is some weighted av-

⁴Such a narrative is also consistent with the foundation of a sovereign wealth fund and the implicit intergenerational distribution of resource wealth which benefits future generations.

⁵The Easterlin Paradox refers to a surprisingly small correlation between reported subjective-well being and national income across countries (Easterlin 1974; Oswald 1997).

erage of the income of the J "relevant others"⁶. This term captures the idea that it is important how much an individual is capable of affording in comparison with the relevant others (Veblen 1899; Duesenberry 1949; Easterlin 1995; Luttmer 2005; Ferrer-i Carbonell 2005). Hence, an increase in average current income of the relevant others affects an individual's satisfaction with income negatively. The last term reflects expectations of current income y_t which were formed x periods ago. This term captures the idea that people may remain dissatisfied with an increase in income if it falls short of their expectations (Ross 2007; Mas 2006; McBride 2010). In recent years several authors have provided evidence that expectations matter in determining the level of satisfaction with income. See in particular Mas (2006) for a field experiment and McBride (2010) for a laboratory experiment.⁷ Here we are emphasizing the last channel as being particularly important during resource booms. Excessive expectations occur particularly often in rapidly changing economic environments as these are easily and quickly adjusted (Ray 2006).

The remainder of the paper is structured as follows. Section 2 provides some background information on the oil boom and the conflicts in Kazakhstan. A simple theoretical framework is presented in section 3. In section 4 a description of the data is provided. Our identification and estimation strategy is discussed in section 5. In section 6 we present our results which are followed by a discussion in section 7. In section 8 we conclude.

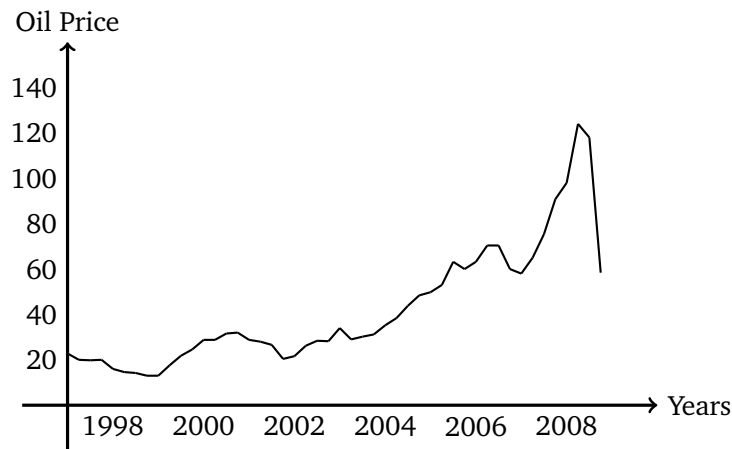
3.2 Oil Boom and Conflicts

Since the dissolution of the Soviet Union in 1991 until 1999 economic growth in Kazakhstan has been dominantly negative or marginally positive (1-2%). The poor economic performance is usually attributed to the transition from a planned to a market economy (Pomfret 2005). Additionally the exogenous development of commodity prices did not allow the resource rich

⁶Relevant others refers to the fact that individuals usually choose whom they compare themselves with. Typically, they do not choose members of the Royal Family or celebrities. Instead, the relevant others tend to be family members, neighbours or colleagues (Frey 2008).

⁷See also Rizzo and Zeckhauser (2003), Crawford and Meng (2011), Abeler et al. (2011), Gill and Prowse (2012) for empirical work on reference dependent preferences.

Figure 3.1: Quarterly Oil Price in US\$.



Notes: Oil Price per Barrel (Cushing, OK WTI Spot Price FOB) (EIA 2012)

country to benefit from its natural resource abundance. This changed in 1999. As shown in Figure 3.1, the oil price started to increase from 12 US\$ per barrel at the beginning of 1999 and had roughly tripled by the end of 2003 and quintupled by the end of 2005. The increase stimulated the development of new oil and gas fields and, eventually, production. A high oil price is particularly important for oil and gas companies in Kazakhstan due to high transport costs driven by its geographic location⁸ and the harsh climate conditions making the extraction of oil and gas particularly costly. Overall, the volume of oil and gas produced in the oil rich districts of Kazakhstan doubled (from 30000 to 60000 thousand tonnes) between 1999 and 2005 (SARK 2011).

GDP per capita in Kazakhstan increased from 1130 US\$ in 1999 up to over 9136 US\$ in 2010. This increase in GDP per capita was mainly driven by the oil and gas rich Oblasts⁹ (henceforth Oil Rich Provinces or ORP), the only provinces (except the capitals) to experience a more than tenfold increase in GDP per capita (SARK 2011). ORPs are concentrated in the western part of Kazakhstan in the north east of the Caspian Sea.

The recent economic development suggests that the regional population might have benefited from the oil boom. But, at the same time, ORPs experienced severe labour conflicts in

⁸Kazakhstan is the largest landlocked country.

⁹First level of local government.

Table 3.1: Conflicts across major oil and gas producers

Consortium	Proximate cause of (violent) conflict	Form of Conflict
Tengizchevroil LLP	2004-2005: Low wages; Bad work conditions; Foreigners are treated better.	Large-scale strikes and affrays.
Petrokazakhstan	2005-2007: Unpaid bonuses; Foreigners are treated better.	Peaceful protest.
Kazmunaygas JSC and Karazhanbasmunay	2010-2011: Unpaid bonuses; Low wages; Bad work conditions.	Large-scale strikes, Deadly riots.
Karachaganak Pet. Oper. B.V.	2001: Low wages; Bad work conditions. 2007: Low wages; Foreigners are treated better and get higher wages.	Peaceful protest.
Mangistaunaygas	2006: Low Wages.	Strikes, Riots.
CNPC - Actobmunaygas JSC	2005: Low Wages; Foreigner are treated better. 2011: Low Wages.	Peaceful protest.

Sources : Yessenova (2008), Shubin (2005) and online newspapers.

the last decades (see Table 3.1). Most notably, the conflict at Tengiz with massive strikes and affrays in 2004-2005, and the recent deadly riots in Zhanaozen in December 2011 in which the local population was involved. The recurrent violent conflicts reveal some dissatisfaction. Table 3.1 is not an exhaustive enumeration, but serves rather as an illustration that conflicts occurred across provinces and consortia¹⁰ (public and private companies). The second column of Table 3.1 suggests that the main reasons for conflicts appear to be low wages and unpaid bonuses, indicating some dissatisfaction with income. However, anecdotal evidence and data from the Statistical Agency of the Republic of Kazakhstan (SARK) suggests that workers in the oil rich region are among the best paid in the whole country.

It is important to emphasize that we are not focusing our analysis on those directly employed in the oil and gas sector. Our focus is on the identification of effects of oil price fluctuations on income satisfaction of households located in proximity to oil and gas fields. Households which might feel entitled to benefit from an increase in the oil price. Before the seminal work of Sachs and Warner (1995), it appeared to be common knowledge that an increase in the price of a natural resource increases the owner's wealth. Economic reasoning suggests that the abundance of an otherwise scarce factor of production should generate profits to the owner. Thus, locals' claims regarding the ownership of natural resource wealth may trigger collective expectations of higher income and conflicts if expectations are not fulfilled.

¹⁰Consortium: a partnership of more than one individual, firm or government (or any combination of these) with the mission to achieve a common goal by participating in a common activity and pooling resources.

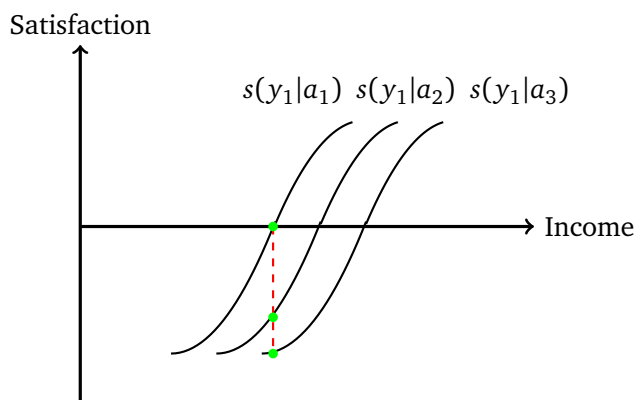
3.3 Theoretical Framework

Standard economic theory suggests that the satisfaction a household derives from income, $s(y_t)$, should increase if income increases over time $t=\{1, 2\}$ ($s(y_2) > s(y_1)$ if $y_2 > y_1$). Aspirations (a_t) might preclude a household deriving more satisfaction from an increased level of income. If households assess their satisfaction with income conditional on the level of aspirations, they might remain dissatisfied if the increase in aspirations exceeds the increase in income ($s(y_2|a_2) < s(y_1|a_1)$ if $y_2 > y_1$ but $a_2 \gg a_1$). As discussed above, previous studies have suggested that past income, relative income and expectations regarding future income seem to be very important in determining the level of aspirations: $a_t = f(\sum_{k=1}^K y_{t-k}\gamma_k, \sum_{j=1}^J y_j\omega_j, E_{t-x}[y_t])$. The first term is some weighted average of income y over the last K periods capturing the effect of past income on aspirations. The second term is some weighted average of the income of the J relevant others. This captures the idea that it is important how much an individual is capable of affording in comparison with the relevant others. The last term reflects expectations of current income y_t formed x periods ago. Now assume that expectations are a function of the publicly observable oil price (p_t): the price of the essentially only locally produced tradable good in the oil rich districts of Kazakhstan. If we are willing to make this assumption, the exogenous increase in the oil price will raise expectations and trigger a shift in aspirations ($a_2(p_2) > a_1(p_1)$ if $p_2 > p_1$).¹¹

Aspirations might in particular pose problems in a quickly changing economic environment, e.g. a resource boom. Aspirations are quite quickly adjusted on the household level. On the other hand, it takes some time until the increased revenues from the oil and gas sector trickle down to the local population via linkages to the local economy. This implies that the increase in the oil price in period t affects aspirations in period t , but the income realized is only affected in a subsequent period x ($s(y_t(p_{t-x})|a_t(p_t))$). Thus, during an

¹¹Anecdotal evidence suggests that this appears to be the case in Kazakhstan (Shubin 2005): The workers logically consider that in conditions in which the cost of oil all over the world has reached record levels, and Petrokazakhstan Oil Products last year processed 30,5 million barrels of oil [...] it is possible to pay people better.

Figure 3.2: Income Satisfaction



Notes: Fixing income in the short run and allowing aspirations to be a positive function of the oil price $a_x(p_x)$, satisfaction with income will decrease if the the oil price increases.

oil price boom the local population might remain dissatisfied for the duration of the boom, $t=\{1, \dots, T\}$, if increasing expectations continually precede the increase in income realized $(\int_{t=1}^T s(y_t(p_{t-x})|a_t(p_t))dt)$. This idea is depicted in Figure 3.2. Hereby, it is assumed that household income is not affected in the short run (y_1) but that aspirations are (a_t). As a result of the continuous dissatisfaction, collectivization and conflicts are more likely to occur (Tocqueville 1986; Gurr 1970; Ray 2006).

The hypothesis to test in the subsequent empirical analysis is:

An increase in the oil price negatively affects households' satisfaction with income.

3.4 Data

The data employed in the analysis is from the Household Budget Survey of Kazakhstan (2001-2005). The survey was created by the SARK in cooperation with the World Bank (World Bank 2004). Since it has been conducted on a quarterly basis. IMF (2003) positively evaluated the survey in 2003. The questionnaire contains detailed information on total household income and expenditure. Of particular interest to the study is that the questionnaire contains questions regarding satisfaction with household income (assessed on a scale from 1 to 5) and the minimum household income necessary not to feel poor as reported by the household head.

The data used is a stratified multi stage¹² cluster sample consisting of 12000 households in each round, representing 0.3 percent of the total number of households (IMF 2003; World Bank 2004). Our focus on the oil rich districts of Kazakhstan reduces our sample to 1800 households. The household sample has been selected from a household register which is based on the 1999 population census. The survey is considered to be representative at the province level.

Officially, the data set is categorized as a repeated cross section but it is a perfectly balanced panel within one year. Approximately 25% of the households are replaced by new households annually. Thus, some households are surveyed for only one year whereas other households remained in the survey for more than 4 years. The introduced rotation still allows us to construct an unbalanced panel across years by identifying households who remain in the sample. The household ID does not change as long as the household remains part of the survey. Once a household is dropped from the survey the ID is assigned to another household which is newly added to the survey. Thus, we used household characteristics which are unlikely to change to identify the panel dimension of the household. In particular, we used household ID, dwelling type, size of the dwelling, the year the house was built, the sex, date of birth and ethnicity of the household head.

Some information has been missing on the reported level of income necessary not to feel poor. To resolve this problem, 120 values have been imputed by using the arithmetic mean of the available observations and 4 observations were dropped. 70 observations were missing on reported satisfaction with income. These values have been imputed by taking arithmetic means of the values which were available for the same household within the same year. Our results are robust to dropping these observations. 52 observations are lost when log-transforming real total household income.

¹²In the first stage, each province (except Almaty and Astana) has been divided into 4 strata (large cities, medium, small towns, and rural settlements). In the second stage, primary sampling units (with at least 150 households) have been chosen. Within each primary sampling unit, households were sampled with sampling probability proportional to the size of the household, and thirty households were listed (10 additional households were listed as replacements).

Using weights neither affects the consistent estimation of the coefficients nor does it offer any significant efficiency gains. Thus, we prefer to use the unweighted specification (Solon et al. 2013) . Weighted means for the variables employed and weighted regression results of our main specification are presented in Tables 3.4 - 3.5 and Table 3.9 (see Appendix), respectively.

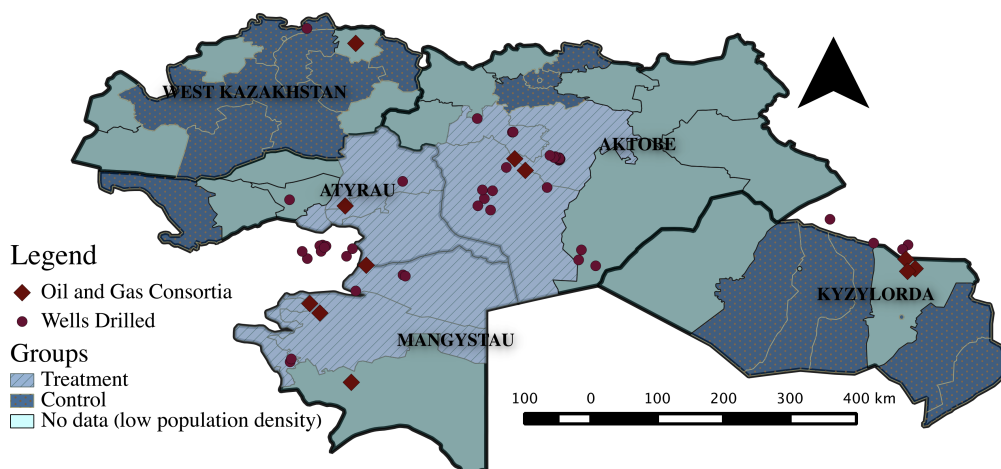
Main variables are presented in Tables 3.4 - 3.5 in the Appendix. The data on provincial inflation has been taken from the SARK (2011).¹³ The geometric mean of yearly inflation is used to calculate real values (discounted to the first quarter of 2001). The exchange rate at the end of the year 2000 was 144.5 Tenge/US\$, such that average quarterly real household income in the sample is slightly above 600 US\$ in the oil rich districts and slightly below 380US\$ in the oil poor districts.

For the spatial identification of the resource rich districts, we construct a data set with the location of oil and gas fields which are operated by the biggest consortia (producing in total c. 95% of oil and gas in 2004) in Kazakhstan. We spatially link the location of the oil fields to the location of households on the district level. Unfortunately, the SARK was not willing to share the precise location of households. Information on the consortia, most importantly their location, is provided in Tables 3.6 - 3.7 in the Appendix and Figure 3.3.

The data for the oil price was taken from EIA (2012) (see Figure 3.1). The oil price is reported on a monthly basis in US\$ per barrel for West Texas Intermediate (Spot Price FOB). The quarterly oil price is calculated by taking the arithmetic mean of the monthly oil prices. The oil price has been transformed into logged values and then first differenced for the empirical analysis.

¹³The SARK uses the Household Budget Survey employed in the subsequent empirical analysis to calculate provincial and national yearly inflation (IMF 2003).

Figure 3.3: Oil rich provinces of Kazakhstan



Notes: The locations of the major consortia listed in Tables 3.6 and 3.7 (see Appendix) have been used to identify treatment and control groups. Consortium: a partnership of more than one individual, firm or government (or any combination of these) with the mission to achieve a common goal by participating in a common activity and pooling resources.

3.5 Identification and Estimation Strategy

The identification strategy employed is basically a difference-in-differences. To identify the effects of the oil boom on household satisfaction with income and household income, two sources of variations are used. First, variation in the oil price: Using the oil price as the treatment has two important advantages in comparison to oil production or revenues. The oil price may be considered to be exogenous to a small open economy. Moreover, the oil price appears to be more likely to affect the formation of locals' aspirations since the increase in oil prices is usually easily observed, as opposed to production and revenues of oil companies. Second, the geographical distribution of oil fields across districts: Following our identification strategy we focus on the five ORPs which account for 100% of the oil and gas production in Kazakhstan. We exclude the remaining provinces from the empirical analysis due to the large heterogeneity across provinces introduced by differences in climate, population density, ethnic diversity and economic structure.¹⁴ The data sample of the ORPs consists of 22 Rayons¹⁵

¹⁴One could argue that fixed effects could be used to account for these differences. However, fixed effects do not account for different trends which might result from differences in initial conditions and thus violate our identification strategy. We have repeated our analysis using the rest of Kazakhstan as a control group and our results do not differ significantly.

¹⁵Second level of local government.

(henceforth district) and 6 cities¹⁶. Districts are defined as being oil rich if an oil or gas field which is operated by one of the biggest consortia (producing in total c. 95% of oil and gas in 2004) is located within a district. The locations of the major consortia listed in Tables 3.6 and 3.7 (see Appendix) were used to identify treatment and control groups (see Figure 3.3). A disadvantage of our identification strategy is that we cannot assign the location of oil fields to individual cities because oil and gas fields are typically not located close to cities. Nevertheless, cities significantly benefit from the oil boom due to the creation of linkages and spending effects. Thus, we add Atyrau and Aktau, the two cities which are located in the core of the oil rich districts to our treatment group. There is plenty of evidence that both cities significantly benefit from the oil boom (Auty 2008; Pomfret 2005). Dropping the cities does not significantly affect our results. We refer to 3 of the cities and 8 of the districts, which are oil and /or gas rich, as the oil rich districts and we refer to the remaining cities and districts as the oil poor districts.

Household fixed effects are employed in the case of the fixed effects (FE) specification, and district fixed effects are employed at the district level in the case of the pooled OLS estimation to account for household specific constant unobservables and for district specific constant unobservables, respectively. All specifications have been estimated using district specific or household specific effects. Quarter fixed effects are included to account for shocks which affect oil poor and oil rich districts simultaneously. The implicit identification strategy in the difference-in-differences framework is that, in the absence of an oil boom, the development of the outcome variable would be similar in oil rich and oil poor districts and the coefficient of the treatment variable would remain insignificant. Hence, the similarity of trends in both areas in the absence of an oil boom is a necessary condition for the validity of the identification strategy. The identification strategy would be invalidated in case of other unobserved time-varying factors correlated with the oil price and differently affecting the oil rich and the oil poor districts. To account for province and district specific time varying unobservables, non-linear district specific or province specific trends are employed in the empirical analysis.

¹⁶As defined by the Statistical Agency of Kazakhstan.

As the treatment in our case is continuous, a distributed lag structure is employed in the empirical analysis for which the transformation and the number of lags is determined theoretically. First, we prefer to use changes in the logged oil price as our treatment. The reason is that we suspect that individuals are not able to recall the level of the oil price. But usually everybody knows whether the oil price has been increasing or decreasing in the recent past. Second, it is well known in the psychological and economic literature that individuals are subject to recall biases (Robinson and Clore 2002). Thus, we suspect that individuals will not be able to remember the change in the oil price from more than a year ago. In particular, up to 4 lags of the oil price are allowed to affect satisfaction with income. The use of changes rather than levels and the choice of lags does not significantly affect our results.

Several additional issues have to be taken into account when working with data on reported satisfaction. First, the use of self-reported satisfaction in the empirical analysis raises the question of the validity and reliability of the results. This problem has been extensively studied in the literature, with the general conclusion that such measures contain a substantial amount of valid variation (Diener 1984; Kahneman and Krueger 2006; Di Tella and MacCulloch 2006).

Second, usually reported life satisfaction (not satisfaction with income) is regressed on income and other covariates expected to affect individual satisfaction (Frey 2008; Kalyuzhnova and Kambhampati 2008; Guriev and Zhuravskaya 2009). Even though insightful, the difficulty of such an approach is the almost infinite amount of variables determining individual satisfaction and, hence, the comparability of such measures (Easterlin 2001). This problem is mitigated because of the way the question is posed in the questionnaire: *"Please advise how satisfied you are with the monetary income of your family within the last three months"*.

Third, prior work on satisfaction identifies the presence of significant individual effects due to fixed personality traits (Frijters et al. 2004). Usually, personal traits are captured using FE estimations. Using FE in a linear probability model implies a cardinalisation of the reported satisfaction. Treating the responses ordinally would require the use of an ordered probit/logit

model. But using FE in an ordered logit/probit framework yields inconsistent estimates due to the incidental parameter problem (Nexman and Scott 1984). The results of a parametric ordered probit estimation without FE imply assumptions which are barely justifiable. In such a context it is assumed that the error term is normally distributed and that there are no household specific unobservables in the error which are correlated with the explanatory variables. This is unlikely to be the case. On the other hand, in the case of an FE estimation, household specific unobservables can be accounted for and the error term is only assumed to be uncorrelated with the explanatory variables, but a cardinalisation of the dependent variable is assumed. In order to solve this problem, we rely on the work of Ferrer-I-Carbonel and Frijters (2004). They conclude that the FE method is the best approach to analyze happiness data. They argue that linear probability model results and ordered probit/logit results are largely consistent with each other, whereas in linear models it is easier to control for fixed effects and interpret the results.

Based on the discussion in the theoretical section, we expect the increase in the oil price to affect satisfaction with income negatively. To formally evaluate the effect, the following regression has been estimated in the case of the FE specification:

$$s_{h,d,t} = \sum_{0=k}^4 \beta_k \Delta \ln(OP_{t-k}) D_d + x'_{h,d,t} \zeta + i'_{h,d,t} \xi + \alpha_h + \gamma_t + \epsilon_{h,d,t} \quad (3.1)$$

$s_{h,d,t}$ is a household's satisfaction with income as reported by the household head for household h , in district d and period t , measured on a scale from 1="not satisfied at all" to 5="fully satisfied"; OP_{t-k} is the lagged quarterly oil price in US\$; D_d is a dummy which is equal to 1 if the household is based in the oil rich districts; $x_{h,d,t}$ is a vector of household specific characteristics controlling for observables¹⁷: education, sex, age, age squared, ethnicity and sector of occupation (public, private, agriculture) of the household head, household size, household size squared; $i_{h,d,t}$ is a vector of income related controls including logged real

¹⁷Household specific controls have not always been available at the quarterly level. Thus, control variables have been used at the yearly level. This does not appear to be a serious drawback as most of the variation takes place on the yearly level.

household income, logged district specific non-self average household income and logged minimum income necessary not to feel poor as reported by the household head. The main purpose of these controls is to account for the alternative channels affecting satisfaction with household income (household income, past income, contemporaneous relative income). The individual channels have been discussed above; α_h is the household fixed effect; γ_t is the time fixed effect and $\epsilon_{h,d,t}$ is the household specific error. Assuming that errors may be correlated over time within a household and across households from the same district, we cluster on the district level (Angrist and Pischke 2009).

In the framework of a natural experiment, we are concerned about the comparability of the oil rich and the oil poor districts. Above, we have argued that by focusing on the five oil and gas rich provinces in Kazakhstan we reduce the heterogeneity of households. But a significant heterogeneity remains between the two groups as can be observed in Tables 3.4 and 3.5 (see Appendix). In the specification above we control for constant unobservables by using household fixed effects and for time varying unobservables on the province level and the district level by using non-linear trends. Nevertheless, we are concerned that the increase in the oil price will differently affect the change in satisfaction due to observable and unobservable household characteristics. To explicitly¹⁸ account for this we reestimate the above specification but use growth in satisfaction rather than satisfaction as our dependent variable:

$$\Delta s_{h,d,t} = \sum_{0=k}^4 \beta_k \Delta \ln(OP_{t-k}) D_d + x'_{h,d,t-1} \zeta + i'_{h,d,t-1} \xi + \alpha_h + \gamma_t + \epsilon_{h,d,t} \quad (3.2)$$

Our specification on the right hand side of the equation is not affected by the transformation, but the interpretation of the variables and their coefficients changes. In particular, the time fixed effect γ_t represents a common shock to changes in satisfaction rather than changes in the level of satisfaction. And the household fixed effect α_h represents a constant growth rate of household satisfaction with income.

¹⁸Implicitly we account for this to a certain degree by using province and district specific non-linear trends.

To test for the implicit assumption that real household income is not affected in the short run, the following specification is estimated:

$$\ln(y_{h,d,t}) = \sum_{0=k}^4 \beta_k \Delta \ln(OP_{t-k}) D_d + x'_{h,d,t} \zeta + \alpha_h + \gamma_t + \epsilon_{h,d,t} \quad (3.3)$$

The specification is similar to equation 3.1 above, except that the dependent variable, $y_{h,d,t}$, is real household income and, $i_{h,d,t}$, the vector of income related controls, is not included in the regression.

Finally, two more general points have to be discussed. As usual, the panel structure is subject to attrition. Attrition in our sample is mainly driven by those households which refused to participate in the survey or move away and are not tracked. Every year approximately 3% of all households have stopped reporting either due to refusal to participate or migration. The effect of attrition is mitigated by construction due to the yearly rotation of households. Whereas the rotation of households mitigates the effects of attrition it also introduces another problem. Our identification strategy is basically a difference-in-differences comparing households in the oil rich districts with households outside the oil rich districts. But households have the possibility to migrate and select into treatment: migrate into the oil rich districts. Thus, the rotation of households introduces contamination, which is an obvious problem in our analysis. In particular, aspirations of migrants are expected to be less affected by the increase in the oil price in comparison to the indigenous population. This is because households which have been based in the oil rich districts before the beginning of the oil boom may feel more entitled to benefit from the oil price boom. To resolve this problem we conduct our analysis using different samples: The complete sample in which attrition is arguably a less serious problem due to the introduced rotation of households; this sample is contaminated by migrants. A subsample of only those households which were part of the survey since 2001; this subsample is not contaminated by migrants. Due to the reasoning presented above, we expect the estimated effect of the former being a downward biased version of the latter.

A similar problem arises due to the fact that the survey is conducted with households being residents in an individual house or part of a house. Foreign or national employees from other provinces who are employed by companies in the oil sector often work in shifts.¹⁹ Due to the construction of the survey, these workers are not captured. This results in an under-sampling of those individuals and households who are permanently located abroad or in other provinces. An important source of dissatisfaction and conflict appears to be the presence of foreigners (see Table 3.1) who, according to national employees, receive a higher wage and better treatment. Given that these individuals are not captured in the survey, we cannot control for the inequality of wages among foreign and national workers as a potentially important source of dissatisfaction. This issue will be further discussed in the section on alternative explanations.

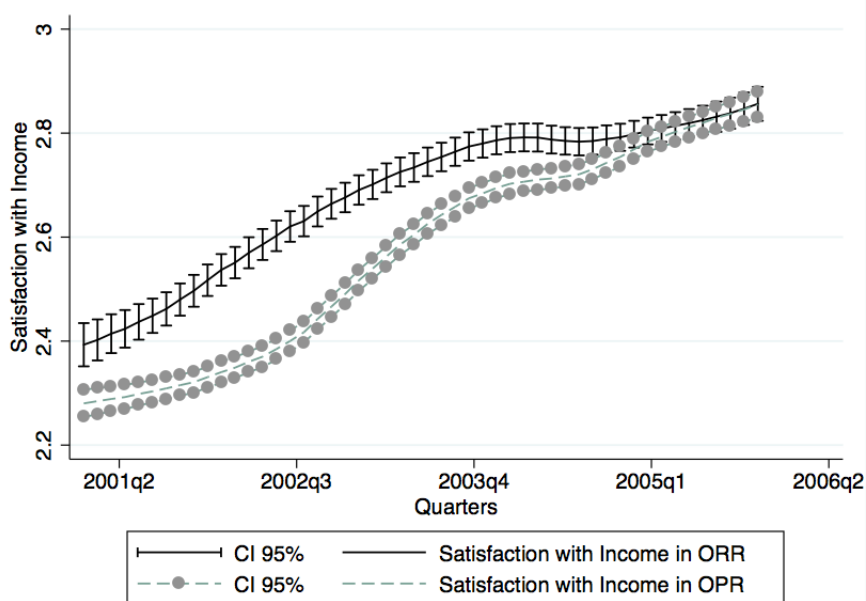
3.6 Results

The development of household satisfaction with income over our sample period is presented in Figure 3.4. Whereas the difference between satisfaction in the oil poor and the oil rich districts is significant at the beginning of our sample, this is not the case at the end of it. According to the previous discussion, this suggests that the treated group experienced either a decrease in income or an increase in aspirations. In Figure 3.5 the development of real household income is presented. Real household income was higher at the beginning of our sample period and in addition increased by a faster rate. Thus, together Figures 3.5 and 3.4 suggest that inflated aspirations might have caused a relative decrease in household satisfaction with income.

The structure of the results presented in Tables 3.8 - 3.10 in the Appendix is identical. In the first column the most parsimonious specification is presented. In columns 2 and 3

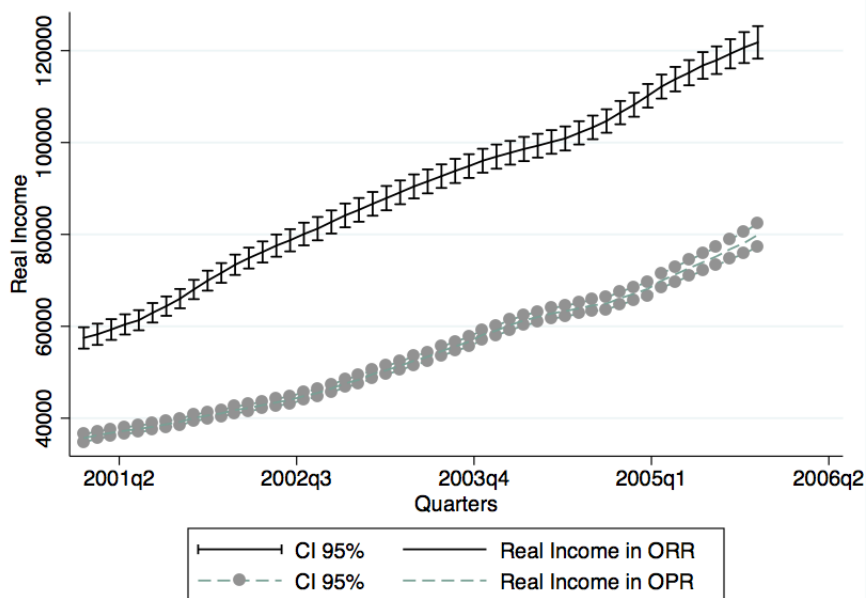
¹⁹Similar to workers on offshore rigs, employees are transported to the oil and gas fields (which are often located in remote areas of Kazakhstan) where they remain for a period of several weeks before they return home.

Figure 3.4: Households' satisfaction with monetary income within the last three months (reported by household head).



Notes: OPD: oil poor districts; ORD: oil rich districts.

Figure 3.5: Total real household income in Tenge



Notes: OPD: oil poor districts; ORD: oil rich districts.

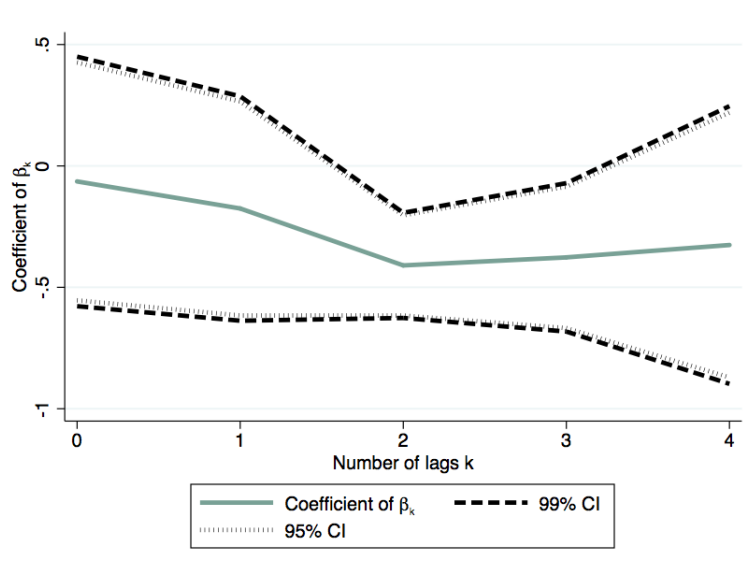
household specific controls and income-related controls are added, respectively. In columns 4 and 5 province-specific trends and province-specific trends squared are added. In columns 6 and 7 district-specific trends and district-specific trends squared are added. Finally, in the

last column an OLS specification is presented where only district-specific effects instead of households fixed effects are employed. In Table 3.8 the results of our preferred specification are presented (equation 3.1). The results of estimating equation 3.1 after weighting the observations according to the sampling strategy are presented in Table 3.9. In Table 3.10 the results of estimating equation 3.1 are presented using just a subset of households. In particular, we use only households which remained within the sample from the beginning of our sample period.

The main result of Tables 3.8 - 3.10 is presented in Figure 3.6, whereas a more detailed summary of the results can be found below. Figure 3.6 represents the effect of oil price growth on satisfaction with income (Column 3 of Table 3.8). The results suggest that satisfaction with income significantly decreases 2 to 3 quarters after the oil price increased. In particular, a 20% increase in the oil prices decreases satisfaction with income by one-tenth of a standard deviation two and three quarters after the oil price increased. The effect vanishes four quarters after the increase in the oil price. This appears intuitive as it takes some time until the change in the oil price is observed and internalized by the local population. Interestingly, the effect of the increase in the oil price is larger for the reduced sample, focusing only on those households who were part of the survey since 2001 (see 3.10 in Appendix). As discussed above, the oil price is arguably not affecting the aspirations of migrants as strongly as the aspirations of those households who have been located in the oil rich districts before the beginning of the oil boom. This is because the latter may feel more entitled to benefit from the boom.

In the previous analysis we have always accounted for the individual channels determining household satisfaction with income by controlling for income related controls (see the theoretical section for a discussion on the individual channels). In what follows, we explore the contribution of the individual channels to satisfaction with income in more detail. In Table 3.11, we successively add income related controls: logged household income (Column 1), logged district specific non-self average household income (Column 2) and the minimum level of household income necessary not to feel poor (Column 3). In Column 4 all the variables are added simultaneously. The effect of logged real income is positive and significant as ex-

Figure 3.6: Coefficients β_k of lag k from estimating specification 3.1



Notes: Number of lag k is depicted on the x-axis. The estimated coefficient β_k of the lagged oil price growth including 95% and 99% confidence intervals are depicted on the y-axis.

pected. In particular, the results suggest that a 25% increase in income increases satisfaction by one tenth of a standard deviation. The non-self district specific average household income affects satisfaction negatively, which is consistent with Easterlin (2001). The results suggest that a 30-40% increase in average income decreases satisfaction with income by one tenth of a standard deviation. Finally, Frey (2008) suggests that reported income necessary not to feel poor is a good indicator for past income. As expected, reported income necessary not to feel poor affects satisfaction with income negatively. Doubling minimum income necessary not to feel poor decreases satisfaction with income by more than one tenth of a standard deviation. Adding average district specific income and minimum income necessary not to feel poor individually, we do not observe the expected negative effect. Coefficients are very small and insignificant. This is most likely due to the omitted variable bias as both indicators are correlated positively with income and, thus, are biased upwards if household income is not controlled for.

In Table 3.12 the results of estimating equation 3.2 are presented. As before, the most parsimonious specification is presented in the first column. In columns 2 and 3 household specific controls and income related controls are added, respectively. In column 4 district

fixed effects are replaced by household fixed effects to allow for a constant household specific growth rate in satisfaction with income. The results do not differ significantly.

Finally, in Table 3.13 the results of estimating equation 3.3 are presented. The results suggest that the increase in the oil price did not have any significant effect on logged real household income within the first year. This is quite intuitive as it takes some time until the positive effects of an increase in the oil price can be observed on the regional level. See Toews (2014) for a detailed discussion of the effects of the oil price boom on employment, household income and income inequality.

Summarizing the results presented in Tables 3.8 - 3.13 (see Appendix) we conclude that the main result is robust to the inclusion of observable household characteristics, household specific constant unobservables and the inclusion of district and province specific non-linear trends. Moreover, the results are robust to different specifications and the use of subsamples. The results are also robust to the exclusion of individual strata and districts.

3.7 Alternative explanations

Financial Flows Abroad:

Enormous net revenues have been generated by the oil sector in Kazakhstan. Net export revenues increased from 2 billion US\$ in 2000 to 35 billion US\$ in 2008. These revenues flowed partly abroad to serve FDI. This might have fuelled locals' dissatisfaction. But the government's share of these massive revenues stagnated around 50% between 2000 and 2008. Moreover, as Esanov and Kuralbayeva (2008) point out, since the beginning of the oil boom the government has successfully undertaken a targeted approach in renegotiating previous arrangements to increase government's take of oil revenues and to reduce the outflow of capital.

Financial flows to central and provincial governments:

The government has put a great emphasize on intergenerational transfers of the oil and gas revenues. For this purpose the National Fund of the Republic of Kazakhstan was founded

Table 3.2: Difference-in-Differences $(\bar{x}_{2005,ORP} - \bar{x}_{2001/2,ORP}) - (\bar{x}_{2005,OPP} - \bar{x}_{2001/2,OPP})$ for self-reported health condition and the level of education.

Variables	$\bar{x}_{2005,ORP}$	$\bar{x}_{2001/2,ORP}$	ORP ⁺ Dif- ference	$\bar{x}_{2005,OPP}$	$\bar{x}_{2001/2,OPP}$	OPP ⁺⁺ Differ- ence	Diff-in- Diff
Health	2.3	2.38 (2002)	-0.08	2.55	2.61 (2002)	-0.06	-0.02
Education	5.14	4.82 (2001)	0.32	5.09	4.77 (2001)	0.32	0

*** p<0.01, ** p<0.05, * p<0.1. Standard Errors are clustered on the district level. +Oil and Gas Rich Provinces, ++Oil and Gas Poor Provinces. The complete sample of 12000 households is used for this analysis.

in 2001.²⁰ Between 2001 and 2008, on average, over 70% of the government's share was saved (Esanov and Kuralbayeva 2008). Also, the ability of subnational governments to capture rents improved over time. Provincial budget per capita was not only higher initially in ORPs in 1999 but also increased by a faster rate. As a result of the relaxed budget constraints, the officially reported level of social expenditures was higher in ORPs (MFRK 2011). Thus, reported monetary flows to provincial and central governments seem to have changed in favour of the local population. On the other hand, the dissatisfaction might originate from the fact that monetary transfers received by local governments disappeared into the pockets of public officials (Caselli and Michaels 2013). Subregional governments are mainly responsible for education and healthcare (World Bank 2005b). Above 50% of local government expenditure goes into healthcare and education. The results of a naive difference-in-differences (with the first /second ²¹ and the last year of the sample) for self-reported health condition and the level of education are presented in Table 3.2. The data on education represents the average level of education on a scale from 1 (no education) -8 (received a university degree). The data on health represents a self-reported condition of health on a scale from 1 (very good) - 5 (very poor). The information on education and health is taken from the Household Budget Survey of Kazakhstan. The results in Table 3.2 suggest that self-reported health and the level of education is better in ORPs. This suggests that even though revenues may have disappeared into the pockets of public officials, health and education, the main targets of provincial expenditure, are unlikely to be the subject of locals' concern. Moreover, we control for non-linear trends on the provincial level in our empirical analysis above. This should account for any

²⁰The main goal of the Fund is to save a portion of the revenues for future generations and to reduce the negative impact of volatile oil and gas prices on the economy.

²¹Data on self-reported health condition was not available for 2001.

Table 3.3: Difference-in-Differences $(\bar{x}_{2005,ORD} - \bar{x}_{2001,ORD}) - (\bar{x}_{2005,OPD} - \bar{x}_{2001,OPD})$ for minimum level of income necessary not to feel poor as reported by the household head.

Variables	$\bar{x}_{2005,ORD}$	$\bar{x}_{2001,ORD}$	ORD ⁺ Differ- ence	$\bar{x}_{2005,OPD}$	$\bar{x}_{2001,OPD}$	OPD ⁺⁺ Differ- ence	Diff-in- Diff
Minimum Income (logged)	12.35	11.75	0.6	11.98	11.38	0.6	0

Notes: *** p<0.01, ** p<0.05, * p<0.1. Standard Errors are clustered on the district level. + oil rich districts, ++oil poor districts.

time varying unobservables on this level.

Employment in the Oil and Gas sector and Migrant Workers:

In the oil rich districts approximately every 5th to every 10th worker of the active labour force is directly employed in the oil and gas sector (Toews 2014). The number of those employed in sectors linked to the oil and gas sector is arguably much larger. However, despite the employment of the local labour force in the oil and gas sector, the presence of migrants in the oil rich districts - which is not captured by the Household Budget Survey and is thus not captured by the relative income indicator (district specific average household income) - is omitted from the empirical analysis. But Table 3.1 suggests that the presence of migrants, with their higher income and their better treatment, is a potentially important source of dissatisfaction in the oil rich districts. However, as presented in Figure 3.6, the negative and significant effect on satisfaction is observed after 2-3 quarters before it vanishes. Such an effect on satisfaction appears inconsistent with the presence of migrants causing dissatisfaction in the oil rich districts. In particular, the increase of the oil price is very unlikely to affect migration, wages, and employment within 2 to 3 quarters. On the other hand, the presence of migrants could be an additional channel in determining individual satisfaction in the long run. Unfortunately, we are not able to test this hypothesis due to a lack of data on the proportion of migrants on the district level. But reported dissatisfaction with migrants appears to be consistent with an argument made by Gurr (1970). He suggests that, once individuals are dissatisfied due to inflated expectations, they become rather rigid in adjusting their expectations downwards and often quick to blame someone else instead, for example, migrants.

Local Inflation

A resource boom is often accompanied by increasing inflation driven by the presence of the spending effect (Cordon and Neary 1982). The basic idea of the spending effect is that the additional money from the windfalls is *inter alia* spent on non-tradable goods creating an upward pressure on local prices. As households in our survey also report how much money they need not to feel poor, we can use this question to test for how households are affected by the resource boom-driven increase in local inflation. As reported in Table 3.3, the results from the simple Difference-in-Differences analysis indicate that no effect can be observed on the amount of money necessary not to feel poor when comparing oil poor and oil rich districts.

3.8 Conclusion

To our knowledge, this paper is the first attempt to use a natural experiment in combination with data on satisfaction with income to provide evidence on the importance of expectations in determining satisfaction. To do this we exploit the exogenous increase in oil prices and our results show that the oil boom in the oil rich districts of Kazakhstan between 2001 and 2005 had a negative effect on household satisfaction with income in the short run. In particular, a 20% increase in the oil prices decreases satisfaction with income by one-tenth of a standard deviation with a lag of two quarters. We argue that this is mainly due to inflated expectations and the upward shift in aspirations. Locals expected to observe an increase in income which did not materialize.

This study has an important policy implication. Our results highlight the relevance of managing expectations in a rapidly changing economic environment. We expect more developing countries to join the group of resource-extracting economies (Ross 2012). These resource discoveries in developing countries are likely to affect the level of development in those countries positively. But, whatever the effect will be, economic adjustments will require time until effects can be observed on the local level. In the meantime, people's expectations will have to be managed in order to avoid increasing dissatisfaction and conflicts.

3.9 Appendix

Table 3.4: Unweighted means, standard deviations and weighted means of the main variables.

Variables	Description	Mean (ORD ⁺)	SD (ORD)	Weighted Mean (ORD)	Mean (OPD ⁺⁺)	SD (ORD)	Weighted Mean (OPD)
Sample Size	Number of Household observations (Cross-section *Time periods)	12276	12276	12276	24140	24140	24140
Real Household Income	Aggregation of quarterly household income from different income sources in Tenge.	89767	76830	88713	54438	50301	53936
Satisfaction with Income	Household satisfaction on a scale from 1-5: Household are asked: Please advise how satisfied you are with your income within the last three months.	2.68	0.9	2.7	2.57	0.91	2.56
Minimum Income necessary not to feel poor	Households have been asked: "Given the current prices what monthly income should a family have with the same number of people as yours not to feel poor?" Reported in Tenge.	203152	147666	199624	140367	94386	138939
District Specific Income	Non-self mean of real household income on the district level in Tenge.	89767	30491	88636	54447	18697	54029
Household Size	Total number of people in a household.	4.74	2.26	4.83	3.96	1.98	3.98
Age	Age of household head.	49	13.5	48	50	13.6	50

Notes: + oil rich districts, ++oil poor districts.

Table 3.5: Unweighted and weighted means of the main dummy variables (Cont. Table 3.4).

Variables	Description	Mean (ORD ⁺)	Weighted Mean (ORD)	Mean (OPD ⁺⁺)	Weighted Mean (OPD)
Dummy Rural Settlement	Dummy is equal to one if the household is located in a rural settlement.	0.29	0.31	0.39	0.42
Dummy city	Dummy is equal to one if the household is located in a large city.	0.47	0.42	0.46	0.42
Dummy for female	Dummy is equal to one if the household head is a female.	0.39	0.38	0.44	0.43
Ethnically Kazakh	Dummy is equal to one if the household head reports to be an ethnic Kazakh.	0.8	0.82	0.66	0.66
High Education	Dummy is equal to one if the household head reports to have graduated.	0.13	0.13	0.2	0.19
Middle Education	Dummy is equal to one if the household head reports to have received a vocational training .	0.47	0.46	0.37	0.37
Lower Education	Dummy is equal to one if household head reports to have a general basic education or general secondary education.	0.34	0.36	0.37	0.38
Employment in the agricultural sector	Dummy equal to one if the household head reports to be employed on a farm or to work on her own farm.	0.04	0.05	0.1	0.1
Employment in the private sector	Dummy equal to one if the household head reports to be employed at a firm, to work for an individual, to work at her own enterprise, to be a member of a production cooperative or to be self employed.	0.26	0.26	0.26	0.25
Employment in the public sector	Dummy equal to one if the household head reports to be employed at a public organization/enterprise.	0.35	0.36	0.28	0.28

Notes: + oil rich districts, ++oil poor districts.

Table 3.6: Accumulative Production of main oil and gas consortia between 2001-2005 (Oil and Gas in 1000 tonnes), ownership and location.

Name	Major Owner	Cum. Prod.	Share (in %)	Province	District	Webpage
Kazakhstan	-	265000	100%	Aktobe, Atyrau, Kyzylorda, Mangystau, West-Kazakhstan	-	-
Tengiz-Chevroil	Chevron, Exxon-Mobil, KazMunayGas, LukArco	65500	25%	Atyrau, Mangystau	Zhylyoyskiy (Atyrau), Beyneuskiy (Mangystau)	www.tengizchevroil.com
EmbaMunay Gas	KazMunay-Gas	13000	5%	Atyrau	Kzylkoginskiy, Isatayiskiy, Makhambetskiy, Makatskiy, Zhylyoyskiy, Atyrau (city)	www.kmg.kz
UzenMunay Gas	KazMunay-Gas	27000	10%	Mangystau	Manghystauskiy, Karakhiyanskiy, Zhanozen (city)	www.kmg.kz
Mangystau-MunayGas	Central Asia Petroleum Ltd. (until 2009)	25000	9%	Mangystau	Manghystauskiy, Karakhiyanskiy, Tupkaraganskiy, Zhanozen (city), Aktau (city)	www.mmg.kz
Karazhanbasmunai	Canada's Nations Energy Ltd. (until 2006)	10000	4%	Mangystau	Tupkaraganskiy, Manghystauskiy	www.kbm.kz
Karashaganak	BP Group, ENI	36000	13%	West-Kazakhstan	Burlinski	www.kpo.kz

Source: Munayshy Public Foundation (2005).

Table 3.7: Accumulative Production of main oil and gas consortia between 2001-2005 (Oil and Gas in 1000 tonnes), ownership and location. (Cont. Table 3.6)

Name	Major Owner	Cum. Prod.	Share (in %)	Province	District	Webpage
Actobe -munaigas	CNPC	24000	9%	Actobe	Temirskiy, galzharskiy, Baiganinskiy	Mu- www.cnpc.ch
Kazakhoil Actobe	KazMunay Gas, LukOil	3500	1%	Actobe	Temirskiy, galzharskiy, Baiganinskiy	Mu- www.koa.kz
Petro - Kazakhstan	Petro - Kazakhstan (until 2005)	23000	9%	Kyzylorda	Syrdarinskiy	www.petro -kazakhstan.kz
Petroleum -Turgai	LukOil, Hurricane Kumkol Munai	13000	5%	Kyzylorda	Syrdarinskiy	www.turgai.kz
KasGer -munai	JSC Yuzh-neftegas, Feba Oil AG, Erbdol Erdgras Gommern	8500	3%	Kyzylorda	Syrdarinskiy	www.kazger -munay.kz

Source: Munayshy Public Foundation (2005).

Table 3.8: Dependent variable: Satisfaction with Household Income

	FE1	FE2	FE3	FE4	FE5	FE6	FE7	OLS
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
β_0	-0.082 (0.255)	-0.073 (0.251)	-0.064 (0.25)	0.049 (0.168)	0.042 (0.17)	0.087 (0.177)	0.072 (0.178)	-.228 (0.365)
β_1	-.166 (0.227)	-.160 (0.226)	-.175 (0.225)	-.057 (0.195)	-.059 (0.196)	-.030 (0.171)	-.031 (0.173)	-.302 (0.295)
β_2	-.395 (0.105)**	-.397 (0.105)**	-.410 (0.105)**	-.346 (0.088)**	-.337 (0.09)**	-.341 (0.104)**	-.331 (0.105)**	-.446 (0.141)**
β_3	-.387 (0.153)**	-.390 (0.152)**	-.377 (0.149)**	-.339 (0.122)**	-.321 (0.134)**	-.320 (0.139)**	-.311 (0.142)**	-.361 (0.165)**
β_4	-.333 (0.28)	-.330 (0.279)	-.326 (0.278)	-.228 (0.213)	-.211 (0.196)	-.195 (0.231)	-.187 (0.203)	-.387 (0.318)
HH ^a FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No
District FE	No	No	No	No	No	No	No	Yes
HH Controls	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Income Controls	No	No	Yes	Yes	Yes	Yes	Yes	Yes
Province Trend	No	No	No	Yes	Yes	No	No	No
Prov. Trend ²	No	No	No	No	Yes	No	No	No
District Trend	No	No	No	No	No	Yes	Yes	No
District Trend ²	No	No	No	No	No	No	Yes	No
Obs.	36416	36416	36364	36364	36364	36364	36364	36364

Notes: Standard errors adjusted for clustering at the district level: *** p<0.01, ** p<0.05, * p<0.1. a Household. Household specific effects or district specific effects, time fixed effects, household size and household size squared are included in every regression. The full set of household characteristics includes the location of the household (big city, small city, rural) and education, age, age², gender, ethnicity and sector of employment (agricultural, public, private) of the household head. The full set of income controls includes real logged household income, district specific non-self average household income (logged), minimum household income necessary not to feel poor as reported by the household head (logged).

Table 3.9: Dependent variable: Satisfaction with Household Income (weighted)

	FE1 (1)	FE2 (2)	FE3 (3)	FE4 (4)	FE5 (5)	FE6 (6)	FE7 (7)	OLS (8)
β_0	-0.079 (0.252)	-0.070 (0.248)	-0.063 (0.248)	0.041 (0.174)	0.035 (0.177)	0.076 (0.184)	0.072 (0.184)	-.221 (0.356)
β_1	-.161 (0.23)	-.156 (0.229)	-.172 (0.228)	-.063 (0.198)	-.066 (0.199)	-.039 (0.175)	-.038 (0.178)	-.296 (0.296)
β_2	-.384 (0.108)***	-.386 (0.108)***	-.401 (0.108)***	-.344 (0.092)***	-.332 (0.094)***	-.340 (0.111)***	-.327 (0.111)***	-.440 (0.139)***
β_3	-.377 (0.148)**	-.380 (0.147)***	-.369 (0.145)**	-.336 (0.124)***	-.311 (0.133)**	-.318 (0.142)**	-.301 (0.142)**	-.361 (0.158)**
β_4	-.332 (0.266)	-.329 (0.264)	-.327 (0.263)	-.237 (0.204)	-.212 (0.191)	-.207 (0.225)	-.185 (0.199)	-.393 (0.3)
HH ^a FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No
District FE	No	No	No	No	No	No	No	Yes
HH Controls	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Income Controls	No	No	Yes	Yes	Yes	Yes	Yes	Yes
Province Trend	No	No	No	Yes	Yes	No	No	No
Prov. Trend ²	No	No	No	No	Yes	No	No	No
District Trend	No	No	No	No	No	Yes	Yes	No
District Trend ²	No	No	No	No	No	No	Yes	No
Obs.	36416	36416	36364	36364	36364	36364	36364	36364

Notes: Standard errors adjusted for clustering at the district level: *** p<0.01, ** p<0.05, * p<0.1. a Household. Household specific effects or district specific effects, time fixed effects, household size and household size squared are included in every regression. The full set of household characteristics includes the location of the household (big city, small city, rural) and education, age, age², gender, ethnicity and sector of employment (agricultural, public, private) of the household head. The full set of income controls includes real logged household income, district specific non-self average household income (logged), minimum household income necessary not to feel poor as reported by the household head (logged).

Table 3.10: Dependent variable: Satisfaction with Household Income (reduced sample)

	FE1	FE2	FE3	FE4	FE5	FE6	FE7	OLS
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
β_0	0.013 (0.314)	0.006 (0.295)	0.009 (0.29)	0.144 (0.228)	0.163 (0.222)	0.162 (0.244)	0.134 (0.238)	-.047 (0.368)
β_1	-.345 (0.306)	-.347 (0.297)	-.350 (0.293)	-.146 (0.234)	-.126 (0.231)	-.123 (0.227)	-.130 (0.224)	-.432 (0.334)
β_2	-.498 (0.147)***	-.495 (0.15)***	-.482 (0.148)***	-.394 (0.144)***	-.338 (0.163)**	-.376 (0.163)**	-.350 (0.171)**	-.513 (0.155)***
β_3	-.438 (0.182)**	-.430 (0.181)**	-.399 (0.177)**	-.387 (0.174)**	-.261 (0.166)	-.358 (0.172)**	-.279 (0.16)*	-.406 (0.189)**
β_4	-.365 (0.331)	-.359 (0.329)	-.338 (0.332)	-.230 (0.314)	-.069 (0.28)	-.187 (0.33)	-.105 (0.302)	-.353 (0.342)
HH ^a FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No
District FE	No	No	No	No	No	No	No	Yes
HH Controls	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Income Controls	No	No	Yes	Yes	Yes	Yes	Yes	Yes
Province Trend	No	No	No	Yes	Yes	No	No	No
Prov. Trend ²	No	No	No	No	Yes	No	No	No
District Trend	No	No	No	No	No	Yes	Yes	No
District Trend ²	No	No	No	No	No	No	Yes	No
Obs.	13928	13928	13891	13891	13891	13891	13891	13891

Notes: Standard errors adjusted for clustering at the district level: *** p<0.01, ** p<0.05, * p<0.1. a Household. Household specific effects *or* district specific effects, time fixed effects, household size and household size squared are included in every regression. The full set of household characteristics includes the location of the household (big city, small city, rural) and education, age, age², gender, ethnicity and sector of employment (agricultural, public, private) of the household head. The full set of income controls includes real logged household income, district specific non-self average household income (logged), minimum household income necessary not to feel poor as reported by the household head (logged).

Table 3.11: Dependent variable: Satisfaction with Household Income

	FE1	FE2	FE3	FE4
	(1)	(2)	(3)	(4)
β_0	-0.048 (0.245)	-0.074 (0.253)	-0.072 (0.251)	-0.064 (0.25)
β_1	-0.160 (0.226)	-0.162 (0.229)	-0.161 (0.226)	-0.175 (0.225)
β_2	-0.407 (0.106)***	-0.398 (0.106)***	-0.398 (0.105)***	-0.410 (0.105)***
β_3	-0.377 (0.151)**	-0.390 (0.152)**	-0.390 (0.152)**	-0.377 (0.149)**
β_4	-0.297 (0.283)	-0.332 (0.281)	-0.330 (0.279)	-0.326 (0.278)
Real Income (logged)	0.341 (0.033)***			0.369 (0.035)***
District Specific Income (logged)		-0.015 (0.084)		-0.245 (0.082)***
Minimum Income necessary not to feel poor (logged)			0.01 (0.029)	-0.058 (0.03)*
HH ^a FE	Yes	Yes	Yes	Yes
HH Controls	Yes	Yes	Yes	Yes
Obs.	36364	36416	36416	36364

Notes: Standard errors adjusted for clustering at the district level: *** p<0.01, ** p<0.05, * p<0.1. a Household. Household specific effects or district specific effects, time fixed effects, household size and household size squared are included in every regression. The full set of household characteristics includes the location of the household (big city, small city, rural) and education, age, age², gender, ethnicity and sector of employment (agricultural, public, private) of the household head. The full set of income controls includes real logged household income, district specific non-self average household income (logged), minimum household income necessary not to feel poor as reported by the household head (logged).

Table 3.12: Dependent variable: Growth in Satisfaction with Household Income (dependent variable in First Differences)

	OLS1	OLS2	OLS3	FE1
	(1)	(2)	(3)	(4)
β_0	0.105 (0.17)	0.108 (0.17)	0.126 (0.177)	0.078 (0.192)
β_1	-.113 (0.197)	-.110 (0.198)	-.090 (0.2)	-.155 (0.22)
β_2	-.358 (0.165)**	-.358 (0.165)**	-.358 (0.165)**	-.392 (0.177)**
β_3	0.089 (0.112)	0.088 (0.112)	0.087 (0.11)	0.069 (0.113)
β_4	0.245 (0.212)	0.246 (0.213)	0.252 (0.212)	0.228 (0.22)
District FE	Yes	Yes	Yes	No
HH ^a FE	No	No	No	Yes
HH Controls	No	Yes	Yes	Yes
Income Controls	No	No	Yes	Yes
Obs.	31528	31528	31477	31476

Notes: Standard errors adjusted for clustering at the district level: *** p<0.01, ** p<0.05, * p<0.1. a Household. Household specific effects *or* district specific effects, time fixed effects, household size and household size squared are included in every regression. The full set of household characteristics includes the location of the household (big city, small city, rural) and education, age, age², gender, ethnicity and sector of employment (agricultural, public, private) of the household head. The full set of income controls includes real logged household income, district specific non-self average household income (logged), minimum household income necessary not to feel poor as reported by the household head (logged).

Table 3.13: Dependent variable: Logged Real Household Income

	FE1	FE2	FE3	FE4	FE5	FE6	OLS
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
β_0	-0.088 (0.117)	-0.088 (0.115)	-0.010 (0.119)	-0.009 (0.117)	-0.056 (0.106)	-0.047 (0.103)	0.087 (0.112)
β_1	-0.010 (0.105)	-0.010 (0.103)	0.081 (0.082)	0.083 (0.081)	0.031 (0.067)	0.038 (0.066)	0.081 (0.113)
β_2	0.035 (0.089)	0.035 (0.09)	0.09 (0.086)	0.11 (0.079)	0.062 (0.091)	0.095 (0.08)	0.068 (0.097)
β_3	-0.027 (0.06)	-0.029 (0.06)	0.013 (0.071)	0.047 (0.056)	-0.011 (0.075)	0.042 (0.055)	-0.047 (0.064)
β_4	-0.091 (0.119)	-0.092 (0.117)	-0.009 (0.107)	0.027 (0.1)	-0.052 (0.088)	0.006 (0.079)	-0.059 (0.115)
HH ^a FE	Yes	Yes	Yes	Yes	Yes	Yes	No
District FE	No	No	No	No	No	No	Yes
HH Controls	No	Yes	Yes	Yes	Yes	Yes	Yes
Province Trend	No	No	Yes	Yes	No	No	No
Province Trend ²	No	No	No	Yes	No	No	No
District Trend	No	No	No	No	Yes	Yes	No
District Trend ²	No	No	No	No	No	Yes	No
Obs.	36364	36364	36364	36364	36364	36364	36364

Notes: Standard errors adjusted for clustering at the district level: *** p<0.01, ** p<0.05, * p<0.1. a Household. Household specific effects or district specific effects, time fixed effects, household size and household size squared are included in every regression. The full set of household characteristics includes the location of the household (big city, small city, rural) and education, age, age², gender, ethnicity and sector of employment (agricultural, public, private) of the household head.

Chapter 4

Getting incentives right: Human Capital Investment and Natural Resource Booms

4.1 Introduction

Accumulation of human capital is widely acknowledged as one of the key preconditions for economic development. A persistent feature of many countries in development traps is low private human capital investment, e.g. Atal et al. (2010). One constraint that private demand for education faces is the lack of resources of the households (Glewwe and Jacoby 2004). However, an equally important problem is that in many cases demand for education is constrained by the lack of job opportunities for the educated labour force. In many African countries, for example, low physical capital investments means that the demand for labour remains low; at the same time, the growth of population leads to a rising labour supply. If the educated labour force grows faster than the overall labour force, the wage for educated workers goes down as well, creating disincentives for acquiring education (Teal 2000). The lack of employment opportunities constrains not only the demand for education, but also the ability of education to contribute to economic growth (Pritchett 2001). However, while the argument seems to be compelling, empirical evidence is still limited, particularly since it is difficult to establish a causal link between the demand for an educated labour force and the

demand for education.

This paper aims to contribute to the literature by explicitly showing how growing employment opportunities result in a growing demand for education in a developing country. To avoid endogeneity, our identification strategy necessary to establish the causal effect relies on looking for an exogenous shock, which influenced the labour market, changing the employment chances for educated labour and, as a response, triggered demand for education. From this point of view, we investigate the oil boom in the first half of the 2000s in Kazakhstan. Kazakhstan is an oil rich country, and the oil price increase of the 2000s had a major influence on its economic development. It is, however, not a country capable of influencing the global oil market; thus, it is safe to say that the rapid increase of oil prices in the first half of the 2000s constituted an exogenous shock for Kazakhstan. At the same time, the demand for an educated labour force from the oil industry of Kazakhstan is substantial due to, first, the technological complexities of oil extraction in harsh climatic conditions in the heart of Eurasia and, second, rigorous local content requirements imposed by the Kazakhstani government on oil companies, which are therefore forced to hire primarily from the local labour market. The identification we use is similar to that of Løken (2010), who, while investigating the impact of Norwegian oil boom, focuses on the effect of growing income on demand for education (while we study the effect of changes in employment opportunities), and Black et al. (2005b), who look at the effect of the coal boom in the US on demand for education in the 1970s.

Our results are based on a unique dataset from the Kazakhstan Household Budget Survey, providing detailed information on the spending and income structure of a panel of Kazakhstani households. We utilize the fact that oil and gas deposits are allocated unequally across the territory of Kazakhstan, providing us with the necessary variation in terms of the effect of a resource boom on different households within the country. We then use difference-in-difference and difference-in-difference-in-difference strategy to identify the effect of the oil boom.

We show that, as a consequence of the oil boom, in oil-rich districts the probability of formal employment for educated labour went up; as a result, there was an increase in private

spending on education (measured by the likelihood of paying tuition fees for tertiary education, i.e. university studies or vocational training). Hence, under certain conditions private demand for education and, thus, the accumulation of human capital might be triggered by a boom in the resource extractive sector. However, we stress that a necessary condition of our result (besides the oil and gas industry being linked to the local labour market) is that the oil boom indeed results in a growing demand for educated labour. Stated otherwise, we argue that in some cases resource booms represent not simply 'manna from heaven', but rather an opportunity for larger revenues, which can be realized only through intensive use of necessary inputs - particularly, educated labour.

The paper is thus related to two literatures. First, our paper is related to the literature investigating the contribution of resource booms to the accumulation of human capital. A well-known result of the resource curse literature is that governments often struggle with an efficient management of resources (van der Ploeg and Venables 2012; Ross 2012). While using the resource revenue to improve the educational level of the population seems to be an attractive option, it is debatable whether governments manage to seize this opportunity. Several studies show that resource abundance causes governments to provide insufficient investments in education (Gylfason 2001; Birdsall et al. 2000; Blanco and Grier 2012; Suslova and Volchkova 2013) or finance education in a very inefficient way (Vicente 2010; Caselli and Michaels 2013). The reason for this outcome is the lack of ability and interest of a rentier state to provide public goods in general. Other papers come to different conclusions: Stijns (2006), Stijns (2009), Pineda and Rodríguez (2010) and Davis (1995) argue that mineral resource wealth may be associated with elevated human capital levels. But independent of how the government performs in terms of education provision, the question remains whether *private* demand for education reacts differently to the resource boom than the public supply. The empirical evidence in this area is limited; most papers investigate developed countries and reach inconclusive results. While some papers argue that changes in the labour market due to the resource boom are likely to reduce the demand for education by private households, at least in the short run (Black et al. 2005a; Walker 2012), others conclude that the resource boom merely delays the timing of schooling but not the long-term outcome (Emery et al.

2011) which may have a positive impact on schooling (Michaels 2011; Braakmann 2011)), no causal impact at all (Løken 2010) or a heterogeneous effect depending on the quality of political institutions (Cabrales and Hauk 2011).

Second, and more generally, we contribute to the investigations concerning the impact of employment opportunities for educated labour on demand for education by offering a novel identification approach and previously unexplored empirical evidence. There exists a growing, but still small literature, focusing on developing countries and showing the positive effects of labour market opportunities on demand for education. Federman and Levine (2004) show that in Indonesia industrialization, leading to the growing demand for an educated labour force, resulted in a growing demand for education. Le Brun et al. (2011) make a similar conclusion for Mexico, while Atkin (2012) suggests that growing employment opportunities in the low-skill sector (e.g. factory openings) resulted in a higher number school dropouts in this country. Heath and Mobarak (2012), on the case of Bangladesh, compare how demand for education reacts on job market opportunities and on direct subsidization of education. They show that a growing demand for education is associated with the former. Oster and Steinberg (2013) argue that in India new jobs originating from an IT service centre in a vicinity of a school increase enrolment in this school. Munshi and Rosenzweig (2006) use the Indian case and show that new jobs, which were accessible to members of various casts, created incentives for low-cast girls to invest more in education; Jensen (2010b), in a field experiment, shows a similar effect of recruitment services for women on educational investments in girls. Still, as early as 1974, Fields (1974) points out that demand for education may remain high even if employment opportunities for educated labour are low.¹

The paper is organized as follows. The next section provides background information on

¹Furthermore, there exists a rich literature mostly for developed countries, which looks at the substitution effects between demand for education and job opportunities: if the latter are sufficiently numerous and the return on education is low enough, the growing job market could crowd out education (Duncan 1965; Neumark and Wascher 1995; Rees and Mocan 1997; Di Pietro 2006). Decline of employment opportunities can (at least temporarily) lead to a growing demand for education as a 'waiting strategy' during business cycles (Betts and McFarland 1995). In developing countries, where the pressure to earn money is higher, the counter-cyclical demand for education is unlikely to be as strong; nevertheless, excessive demand for university education, in an attempt to delay market entry in a hostile environment has been argued to be present in transition countries as well, see Kuzminov (2004).

Kazakhstan. In the third section we present a simple theoretical model, which allows us to derive the main predictions for the empirical analysis. The fourth section describes the data. The fifth section presents the econometric strategy of our analysis. The sixth section summarizes the main results. In the seventh section we discuss some of our concerns. The last section concludes.

4.2 Empirical Case: Kazakhstan

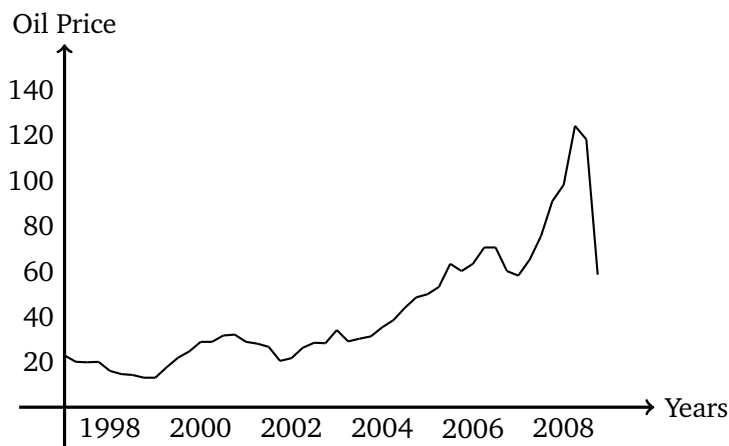
Kazakhstan is one of the largest oil and gas producers in the world, ranking 17th in terms of total oil production in 2011 according to the Energy Information Administration data. In recent years oil and gas accounted for more than 70% of the country's exports. Thus, it is a classical resource-rich country, in which negative effects of resource orientation on public supply of education are possible.

In the early 2000s Kazakhstan experienced an oil price boom, associated with skyrocketing prices on the global oil markets. Figure 4.1 plots the dynamics of the oil prices over the period in question: one can see that since 2001 there was an almost steady increase in oil prices. As a result, investment and employment in Kazakhstan's oil and gas industry grew at a very fast pace Toews (2014). Thus, the resource boom of the 2000s provides the natural experiment necessary for the investigation of our paper.

Oil extraction in Kazakhstan is due to environmental and ecological conditions and the geographical location (far from convenient export ports) much more challenging than in e.g. Gulf countries. The conclusion that Kazakhstan poses a particular technological challenge for oil extraction is widespread (Kaiser and Pulsipher 2007). This is especially true for the major offshore oil fields. Hampton Fowler, Halliburton's production manager, described the conditions at Kashagan² as an unusual combination of "arctic-conditions, high pressure, sour service, zero discharge operations" (Lang 2003). Under these conditions, it is not surprising

²One of the biggest recent offshore discoveries in the world.

Figure 4.1: Quarterly Oil Price in US\$ per Barrel (Cushing, OK WTI Spot Price FOB)



Source: EIA (2012).

that oil extraction relies on a well-trained staff in oil companies. In fact, the lack of qualified personnel is a recognized problem, repeatedly pointed out by the oil companies themselves.³ As a result, it is safe to say that in Kazakhstan there is strong demand for an educated labour force from the oil sector.

The oil sector has a major influence on employment in oil rich districts. According to Toews (2014), in the two main oil rich provinces - Atyrau and Mangistau - every 10th and every 5th worker, respectively, works in the oil and gas sector. Of seven major oil producers, three are Kazakhstani domestic companies mostly employing a local labour force; four are multinationals, with the share of local labour force exceeding 90% for three of them and 80% for one of them.

There are three reasons for this employment pattern. First, Kazakhstan is located in the centre of Eurasia, has very harsh climate conditions⁴ and a relatively strict visa regime,⁵ making foreign labour (except for the neighbouring countries of Central Asia and Russia) relatively expensive to import. Second, the Kazakhstani government requires all oil and gas compa-

³E.g. <http://articles.gazeta.kz/art.asp?aid=57521>, accessed July 30, 2013, or <http://oilkz.kz/ru/news/article/376>, accessed July 30, 2013.

⁴To provide an example, in Atyrau the average monthly temperature varies from -6.4° Celsius to 26.8° Celsius over the year, with extreme temperatures recorded as varying between -37.9 and 42.7° Celsius.

⁵Except for post-Soviet countries, visa-free entry to Kazakhstan for regular travellers is provided only for Mongolia, Turkey, Hong Kong and Serbia.

nies to procure most goods and services through tenders giving preference to local suppliers. The 1995 Petroleum Law required most subcontractors to be Kazakhstani organizations, and the 1996 Law on Subsurface and Subsurface Use set requirements for using local personnel (Ospanova 2012). Thus, the link between the oil sector and the local labour market is protected through 'local content' requirements. Third, oil companies can rely on a relatively well-trained staff and education system inherited from the Soviet period - we will discuss this in what follows.

Kazakhstan inherited a developed educational infrastructure from the Soviet Union, including a network of universities and vocational training facilities in all major cities of the country. In 2005, the gross secondary school enrolment accounted for 95%, and the gross tertiary school enrolment for 53%. It is debatable whether this education system produces workers of sufficient quality, but the situation is definitively much better than in many developing countries. In developing countries local educational systems frequently do not exist and even simple literacy of the labour force is not guaranteed. In Kazakhstan illiteracy is negligible. Actually, in international comparison, Kazakhstan performs quite well. In the most recent Global Competitiveness Report (2013-2014) it scored 54th among 148 countries in terms of 'Higher Education and Training'. Throughout the last five years, Kazakhstan has been consistently included in the top five countries in the world in the UNESCO 'Education for All Development Index'.⁶ Also, the World Bank (2005a) acknowledges that Kazakhstan has a solid base for training of professionals in the oil sector.

In Kazakhstan the number of individuals without secondary education is very small. So we do not obtain the necessary variation in this respect; the state still ensures that almost all children attend a secondary school. The main education decision to be made by households is whether to invest in tertiary education. Limiting our attention to university education would be too restrictive, since an oil boom is likely to have an effect not only on a very small group of employees with university education, but also on specialized blue-collar workers. Tertiary education in Kazakhstan consists of universities and technikums (vocational training educa-

⁶One has to acknowledge that this is to a large extent determined by the quality of primary and secondary education.

tion institutions in the Soviet Union countries providing training for specialized blue-collar jobs based on 3-year programs). Before 2004, admission to technikums and universities was based on entrance exams. These were determined by the educational facilities jointly with governmental agencies. Since 2004, the entrance exams were replaced by the Single National Test, a unified exam sat by all high school graduates in the country. The autonomy of educational institutions in Kazakhstan is limited by constant supervision and control by the governmental agencies; standardized curricula are taught throughout the country.

Education generally requires the payment of tuition fees; there is a system of governmental grants covering costs of education in a variety of disciplines and universities. As of 2011, 77% of university students have had to pay tuition fees. Therefore, and crucially for our research question, the decision to obtain tertiary education in Kazakhstan is costly.

4.3 Theoretical Framework

Based on the discussion above we suggest the following theoretical framework. The representative profit maximizing firm combines capital K , educated and uneducated labour, L_e and L_u respectively, using a constant return to scale production function to extract a natural resource $X_r = q(K, L_e, L_u)$. The good is traded internationally for the price p_r . There is a minimum wage in both labour markets (educated and uneducated) such that $w_e = \bar{w}_e > w_e^*$ and $w_u = \bar{w}_u > w_u^*$, where * indicates the market clearing wage. The assumption of a minimum wage may be debatable for developing countries, where governments typically fail to maintain strict labour market regulations.⁷ But there are several more realistic ways to interpret it. First, it may be the 'socially acceptable' wage which educated and uneducated labour is ready to work for in a particular country due to social status. Second, it may be the wage necessary to cover the exogenous costs of living, which may be relatively high.⁸ The price of the natural resource, the production technology and the minimum wage are determined

⁷In Kazakhstan, a minimal wage imposed by the government has officially existed since the collapse of the Soviet Union; this wage is not always enforced and typically is below the minimal costs of living.

⁸This is very important in the case of Kazakhstan due to its harsh climatic conditions; these make costs of housing (to reach a minimum quality standard), winter clothing, etc. high, while necessary for survival.

exogenously. Firms rent the optimal level of capital from international capital markets and decide on the optimal level of employment such that $r = p_r q'_K$, $w_e = p_r q'_{L_e}$ and $w_u = p_r q'_{L_u}$.

The total number of educated and uneducated workers is given by $N = N_e + N_u$. As a result of the minimum wage, there is involuntary unemployment in both markets. It is assumed that uneducated workers cannot be employed as educated workers. Moreover, we assume that due to better outside options educated workers prefer to remain unemployed if not offered a position which befits their education (Teal 2011). In both labour markets places are attributed randomly. Thus, the probability of being employed as an educated worker and as an uneducated worker are $\frac{L_e}{N_e}$ and $\frac{L_u}{N_u}$ respectively. As in Harris and Todaro (1970), the expected wages of educated and uneducated workers are then given by

$$\begin{aligned}\omega_u &= \bar{w}_u \frac{L_u}{N_u} \\ \omega_e &= \bar{w}_e \frac{L_e}{N_e}\end{aligned}\tag{4.1}$$

By construction the change in expected wages is determined by the probability of being employed in a particular capacity. Note that the relative change in expected wages is driven by the educated labour intensity of the industry and the relative size of the labour markets.

Every household consists of one worker and one child. Individuals live for one period only and have the following utility $u_i = \ln c_{1,i} + \beta \ln c_{2,i}$. $c_{1,i}$ is representing the current consumption of household i . $c_{2,i}$ indicates the child's consumption in the subsequent period which is weighted with β . The inter temporal budget constraint is given by $y_i + \omega_{2,u}/(1+r) + e_i(\omega_{2,e} - \omega_{2,u})/(1+r) - e_i\theta = c_{1,i} + c_{2,i}/(1+r)$. $y_i \in \{\bar{w}_u, \bar{w}_e, w_i\}$ indicates the household income in period 1, whereas w_i represents income from the informal sector. e_i is equal to one if the household decides that the child should receive tertiary education and is zero otherwise. If the child successfully completes her education she is considered to be an educated worker. There is heterogeneity among children in their ability to receive education and it is more costly to educate a less able child. Ability is assumed to be uniformly distributed such that

the costs of receiving education are also uniformly distributed $\theta_i \sim U[0, \Theta]$. Θ indicates the most costly education. Households are perfectly informed about the conditions in the labour market and the ability of their own child such that income maximizing households make their decision based on the following rule:

$$\frac{\omega_{2,s} - \omega_{2,u}}{(1+r)} = \theta^* \quad (4.2)$$

The costs of the marginal child to get education θ^* increases as the expected return from education increases. Denoting the share of children who have an incentive to receive tertiary education by S_{θ^*} and using the assumption of the uniform distribution we get:

$$\frac{\omega_{2,s} - \omega_{2,u}}{(1+r)\Theta} = S_{\theta^*} \quad (4.3)$$

Denoting total absorption $A = L_e \bar{w}_e + L_u \bar{w}_u + (N - L_e - L_u)w_i$ we can derive an expression for the average share of expenditure on education:

$$\frac{1}{A} \int_0^{\theta^*} \frac{1}{\Theta} \theta d\theta \quad (4.4)$$

Taking the first derivative with respect to θ^* gives $\frac{\theta^*}{A\Theta}$ which is positive, suggesting that the share of total expenditure spent on education should increase as the ability of the marginal worker decreases. The intuition of this simple model is straightforward. During an oil price boom p_r increases and affects expected wages of educated and uneducated workers (see equation 4.1). Agents observe the current labour market and adjust their expectations about employment opportunities. According to the expectations formed, agents decide whether to invest in their children's tertiary education. Thus, the hypothesis to test in the subsequent empirical analysis is:

If expected return to tertiary education increases during the resource boom the share of households investing in tertiary education increases.

4.4 Data

The data employed in the analysis is from the Household Budget Survey of Kazakhstan (2001-2005). The survey was created by the SARK in cooperation with the World Bank (World Bank 2004). Since then it has been conducted on a quarterly basis and was positively evaluated by IMF (2003) in 2003. The questionnaire contains detailed information on wages, employment, and, most importantly for the underlying research question, education.

The data is a stratified multi stage cluster sample consisting of 12000 households in each round (IMF 2003; World Bank 2004). The household sample has been selected from a household register, which is based on the 1999 population census. The survey is considered to be representative on provincial and national levels. In the first stage, within each province (except Almaty and Astana), areas have been appointed into 4 strata: large cities, medium cities, small towns, and rural settlements. In the second stage, primary sampling units with at least 150 households have been chosen in each stratum. Within each primary sampling unit, households were sampled with sampling probability proportional to the size of the household and thirty households were listed (10 additional households were listed as replacements). Given the homogeneity in weights, the sample can be considered to be self-weighted. Thus, we prefer to use the unweighted specification and weighted results are available upon request.

Officially the data set is categorized as a repeated cross section. It is a perfectly balanced panel within a year as households are surveyed quarterly. Yearly, approximately 25% of the households are replaced by new households. Thus, some households are surveyed for only one year, whereas others remained in the survey for more than 4 years. The introduced rotation allows us to construct an unbalanced panel across years by identifying households who remain in the sample. The household ID does not change as long as the household remains part of the survey. Once a household is dropped from the survey, the ID is assigned to another household which is newly added to the survey. Hence, we use household characteristics unlikely to change, to identify the panel dimension of the household. In particular, we use household ID, dwelling type, size of the dwelling, the year the house was build, sex, date of

birth and ethnicity of the household head. We conduct the analysis of wages and employment on the quarterly level. Information on education expenditure is collapsed up to the yearly level. This is done to avoid noisy fluctuations of education expenditure within a year.

The panel structure is subject to attrition. Households and individuals which did not participate in the survey for a complete year have been dropped. The complete case method has been chosen because dropping households and individuals resulted in a loss of 3% of the total sample on the household level and 5% on the individual level. Both numbers represent a small share of the sample and thus are quite unlikely to affect the consistency of our estimates.

In Table 4.5 in the Appendix a description of the main variables may be found. Means and standard deviations are presented for the oil rich districts (ORD) and the oil poor districts (OPD) (see next section for a definition of the oil rich districts). The data on provincial inflation has been taken from the SARK (2011). The geometric mean of yearly inflation is used to receive real values. The exchange rate at the end of the year 2000 was 144.5 Tenge/US\$, such that average yearly real household expenditure in the sample is around 2450 US\$ in the oil rich districts and the average real household expenditure is slightly above 1820 US\$ in the oil poor districts.

For the spatial identification of the resource rich districts, we construct a data set with the location of oil and gas fields which are operated by the biggest consortia (producing in total c. 95% of oil and gas in 2004) in Kazakhstan. We spatially link the location of the oil fields to the location of households on the district level. Unfortunately, the SARK was not willing to share the precise location of households. Information on the consortia, most importantly their location, is provided in Tables 4.3 and 4.4 in the Appendix and Figure 4.2.

The data for the oil price was taken from EIA (2012), where the oil price is reported on a monthly basis in US\$ per barrel for West Texas Intermediate (Spot Price FOB). The quarterly oil price is calculated by taking the arithmetic mean of the monthly oil prices. The oil price

has been transformed into logged values for the empirical analysis.⁹

4.5 Identification and estimation

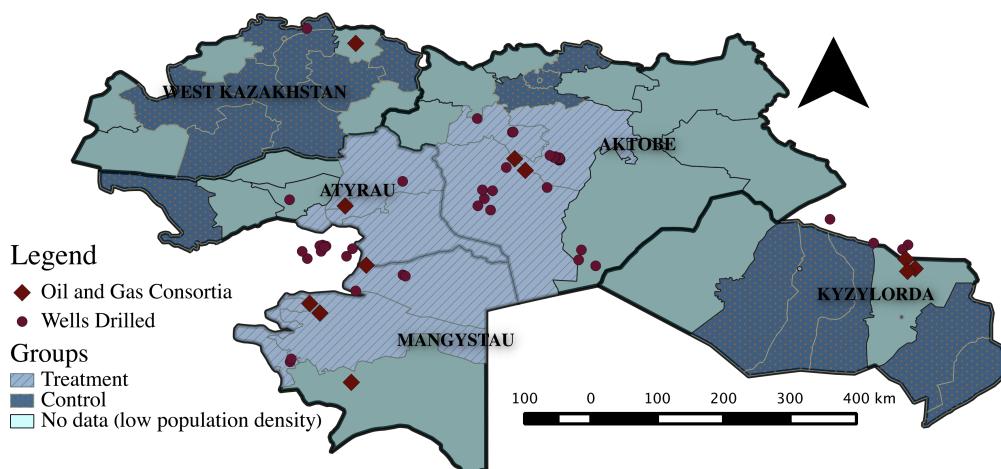
Our identification strategy relies on difference-in-differences or difference-in-difference-in-differences estimators, depending on the hypothesis we test. In our difference-in-differences specification we exploit two types of variation. Firstly, we look at the distribution of oil fields across districts. The treatment group is composed of households living in oil-rich districts; the control group consists of households living in oil-poor districts. Districts are defined as being oil rich if an oil or gas field which is operated by one of the the biggest consortia¹⁰ (producing in total c. 95% of oil and gas in 2004) is located within a district. The locations of the major consortia are listed in Tables 4.3 and 4.4 (see Appendix). Unfortunately, our identification strategy does not allow us to identify cities as being oil rich. This is because cities rarely contain an oil or gas field within their borders. On the other hand, cities in oil rich regions typically significantly benefit from a boom due to developing forward and backward linkages and, more importantly, the spending effect (Cordon and Neary 1982). Thus, we add Atyrau and Aktau, the two cities which are located in the core of the oil rich district to the group of oil rich districts. There is plenty of evidence that both cities significantly benefit from the oil boom (Auty 2008; Pomfret 2005). Dropping the cities does not significantly affect our results.

Figure 4.2 illustrates the allocation of oil poor and oil rich districts (denoted as OPD and ORD henceforth) in the five provinces of Kazakhstan we investigate. We focus on the five provinces in Kazakhstan in which the total amount of oil is produced (see Figure 4.9 in the Appendix). While using sub-national variation is helpful to limit the unobserved heterogeneity (Snyder 2001), focusing on a limited number of provinces located in a particular part of the country allows us to reduce this problem even further by looking at territories with ho-

⁹We understand that part of the oil and gas exported by Kazakhstan is sold to its neighbour countries of the former Soviet Union (especially Russia), with the price level typically lower than on the global markets. However, this price is highly correlated with the price on the global market; thus, essentially, the magnitude of effects we estimate represents the lower bound of the actual effect.

¹⁰Consortium: a partnership of more than one individual, firm or government (or any combination of these) with the mission to achieve a common goal by participating in a common activity and pooling resources.

Figure 4.2: Oil rich provinces of Kazakhstan



Notes: The locations of the major consortia listed in Tables 4.3 and 4.4 (see Appendix) have been used to identify treatment and control groups. Consortium: a partnership of more than one individual, firm or government (or any combination of these) with the mission to achieve a common goal by participating in a common activity and pooling resources.

mogenous climate, environment, ethnic and economic structure.¹¹

Second, the exogenous fluctuations in the oil price affect economic profitability of oil and gas reservoirs, change the level of activity in the oil and gas sector and eventually affect local labour markets in oil rich districts. Kazakhstan is not part of any cartel on the resource markets (like the OPEC), does not dominate the oil markets (even locally, where it has to compete with Russia and other fossil-fuel-rich Central Asian states) and can be essentially considered a price-taker on the global oil markets.

For our diff-in-diff-in-diff specification we require one more source of variation which is the individual's level of education for which we construct a dummy. We define as educated an individual, who possesses some form of tertiary education: either a university degree or vocational training (technikum). The decision to look at the dummy for tuition fees as a dependent variable is driven by several considerations. First, it is more appropriate in terms of our theoretical model, which looks at educational investment as a binary choice. Second, the

¹¹One could argue that fixed effects could be used to account for these differences. However, fixed effects do not account for different trends which might result from differences in initial conditions and thus violate our identification strategy. We have repeated our analysis using the rest of Kazakhstan as a control group and our results do not differ significantly.

information on the exact amount of spending for education is often imprecise, partly because of widespread corruption in the Kazakhstani educational system, but also because it is not easy for households to clearly attribute certain types of expenditure to 'education'. Paying tuition fees, however, clearly indicates that a member of the household is receiving education, and we can avoid these measurement problems.

Regression 4.5 is estimated using quarterly data (the highest frequency data available to us). Model 4.6 is estimated using annual data. This is because tuition fees are typically paid once a year and thus the quarterly variation of this indicator is not informative (typically, it results from differences in the accounting procedures of schools).

We expect that the effect of a resource boom on the local labour market may not be immediate and may be delayed by several quarters or years. For example, enrolment in an educational facility is typically possible only in a particular month (September) and may require specific preparations and a substantial search effort. Wage increases and hiring decisions are also not realized immediately. Thus, we estimate a set of distributed lag models. Including several lags of the oil price directly would result in substantial multicollinearity. Hence, we follow the standard approach of the time series econometrics and transform the equation, transforming all the lags but the last into first differences.¹² This transformation eliminates the problem of multicollinearity and at the same time allows us to interpret the estimated coefficients as an accumulated effect of the oil price on economic outcomes.

To formally test whether returns to education increased during the boom, we estimate equation 4.5 by focusing on two variables affecting returns to education, the real wage and the probability to be employed as an educated or an uneducated worker (see equation 4.1):

¹²In particular, a linear transformation is employed in the following form (where y is a function of x): $y_t = \beta_1 x_t + \beta_2 x_{t-1} + \dots + \beta_{z+1} x_{t-z} = \beta_1 \Delta x_t + (\beta_1 + \beta_2) \Delta x_{t-1} + \dots + (\beta_1 + \beta_2 + \dots + \beta_{z+1}) x_{t-z}$.

$$y_{i,d,t} = \sum_{k=1}^{16} \alpha_k \ln(OP_{t+1-k}) D_d + \sum_{k=1}^{16} \beta_k \ln(OP_{t+1-k}) D_d e_i + x'_{i,d,t} \zeta + \epsilon_{i,d,t} \quad (4.5)$$

$$\epsilon_{i,d,t} = d_d + \gamma_t + e_i + e_i d_d + e_i \gamma_t + v_{i,d,t}$$

$y_{i,d,t}$ is the outcome variable of individual i , in district d in period t and is either the logged real wage of individuals employed in the formal sector or the probability to be employed as an educated worker (see Table 4.5 in the Appendix for a definition of the variables). The latter is captured by a binary variable indicating whether or not an individual is employed in the formal sector (private and public).

The use of these dependent variables requires clarification. Kazakhstan, as many other developing countries, has a large informal employment. As of 2009, roughly one-third of Kazakhstani employment occurred in the informal sector, with 62% of this informal employment being agricultural (for discussion of informal employment see Rutkowski (2011)). For the purpose of this paper, however, we concentrate on employment in the formal sector. This for two reasons: a conceptual one and a data-driven one. First, the wage premium associated with the need to use high-educated labour in oil and oil-related sectors is unlikely to be paid to informal employees. It also holds for the data in our sample: the average wage received in the formal sector is seven times larger than the wage received in the informal sector. In addition, foreign investors and state-owned companies, which dominate the Kazakhstani oil sector, are restricted in their ability to use informal labour, particularly in positions with certain responsibilities (as are most educated labour jobs). It is not surprising therefore that the share of informal employment for professional and technical positions in the total employment is very low (as of 2009, it was 3% and 9% respectively, according to Rutkowski (2011)). Second, while the data we have contains numerous respondents, who identified themselves as being engaged in various forms of informal employment (particularly in the agricultural sector), in the case of Kazakhstan it is very difficult to actually separate informal employment from unemployment. Given the very low level of unemployment benefits, most unemployed are forced to take at least short-term jobs or engage in subsistence agriculture (true even for an

urban population). Thus, we apply a restrictive definition, focusing on formal employment. We look at both public and private employment, because oil extraction in Kazakhstan is partly carried out by the state-owned company, KazMunaiGaz.

OP_{t-x} is the yearly logged oil price in US\$. We allow up to four years for the oil price to trickle down and affect the local labour markets and the individual household decisions; D_s is a dummy variable, which takes the value one if the household is located in an oil rich district and zero otherwise; e_i is a binary variable indicating whether an individual has received a tertiary education or not (as defined above); $x_{i,s,t}$ is a vector of individual specific characteristics controlling for observables: sex, age and ethnicity; d_s are district specific fixed effects; γ_t are quarterly fixed effects to control for common quarterly shocks; $v_{i,s,t}$ is the individual specific error. Assuming that the error may be correlated over time within a household and across households from the same district in the same period, Angrist and Pischke (2009) suggest clustering at the district level, the highest possible level of clustering.

To evaluate our main proposition we estimate the following specification:

$$y_{h,d,t} = \sum_{k=1}^4 \beta_k \ln(OP_{t+1-k}) D_d + x'_{h,s,t} \zeta + \alpha_h + \gamma_t + v_{h,d,t} \quad (4.6)$$

$y_{h,d,t}$ is the outcome variable of household h , in district d , in year t . The outcome variable is a binary variable indicating whether a household reported paying tuition fees for a university or a technikum; the $x_{h,d,t}$ is a vector of household specific characteristics controlling for observables: education and sector of occupation (public, private, agriculture) of the household head, household size and household size squared; α_h is household fixed effect; $v_{h,d,t}$ is the household specific error. As before, we cluster on the district level.

All estimations for binary dependent variables are obtained using linear probability model (LPM). The application of LPM vs. logit or probit in this context is subject to debate; LPM is advocated, for example, by Angrist and Pischke (2009), particularly for the case of large sam-

Table 4.1: Shares of total household expenditure on different types of goods across districts and years.

Variables	ORD ⁺ (2005)	ORD (2001)	ORD Differ- ence	OPD ⁺⁺ (2005)	OPD (2001)	OPD Differ- ence	Diff- in-Diff
Food	44%	59%	-15%	45%	56%	-11%	-4%*
Daily Expenditure (excl. Food)	8%	11%	-3%	11%	10%	-1%	-2%*
Clothing	11%	7%	4%	10%	7%	3%	1%
Durable Expenditure (excl. Clothing)	6%	2%	4%	4%	2%	2%	2%*
Education	6%	3%	3%	4%	2%	2%	1%**
Health	2%	2%	0%	2%	2%	0%	0%
Services	4%	2%	2%	3%	2%	1%	1%**
Transport	3%	3%	0%	3%	3%	0%	1%
Utilities	6%	6%	0%	6%	6%	0%	0%
Fuels	2%	2%	0%	4%	4%	0%	0%
Sum	>90%	>90%		>90%	>90%		

⁺Oil Rich Districts, ⁺⁺Oil Poor Districts.

ples. For us, using LPM is important because our analysis heavily relies on multiple interaction terms and a complex panel data structure, which makes interpretation of marginal effects in logit or probit extremely difficult (on interaction terms in probit or logit see (Ai and Norton 2003)).

Essentially, this means that we estimate two models. In the first model we employ district fixed effects and compare labour market outcomes following the oil price boom on the wage of the educated, and the probability of being employed as an educated worker in comparison to the uneducated. Here our units of observation are *individuals* and we use quarterly data. In our second specification our units of observation are *households* and we use yearly data. Here we test whether households actually increase their expenditure on tertiary education using household fixed effects.

4.6 Results

As a first step in discussing the implications of the resource boom for educational expenditures, we start with a simple comparison of means for resource-rich and resource-poor districts

Figure 4.3: Proportion of HH reporting the payment of tuition fees.

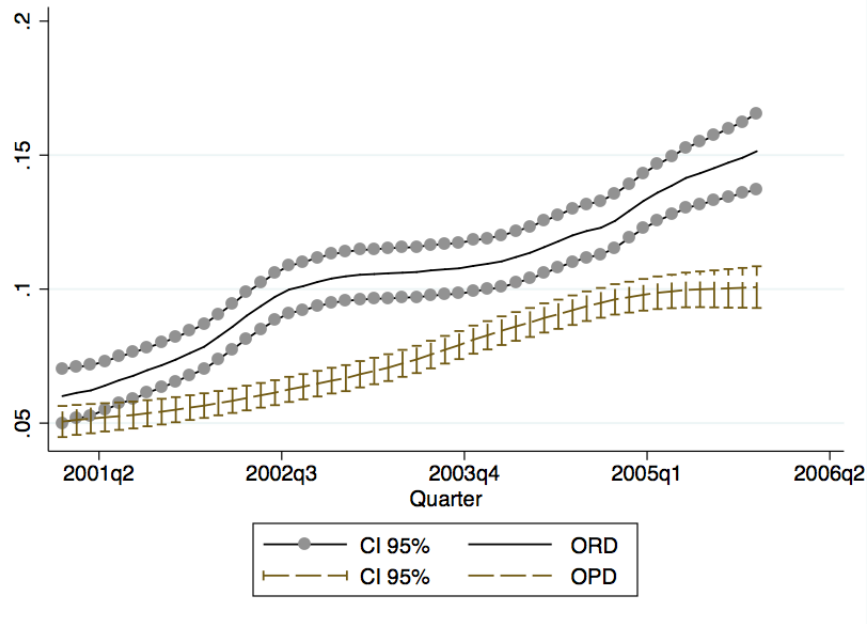
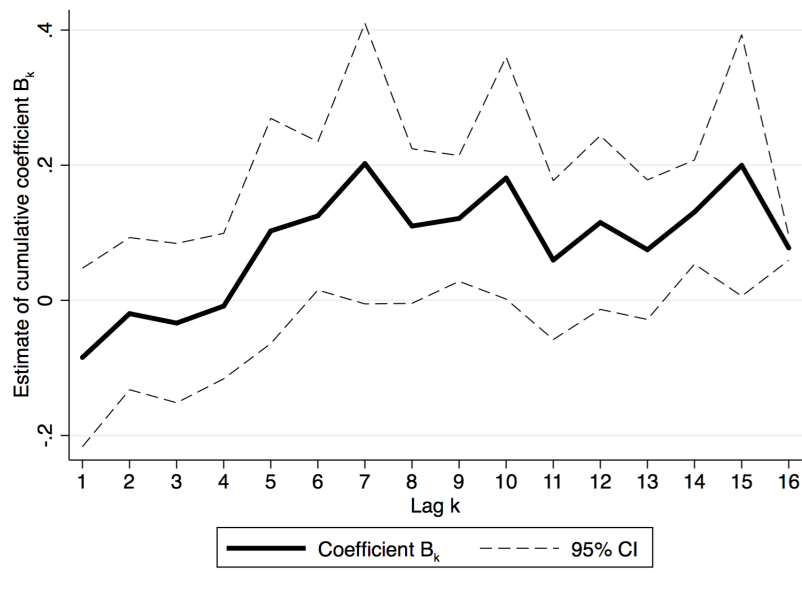


Figure 4.4: Employment in the Formal Sector



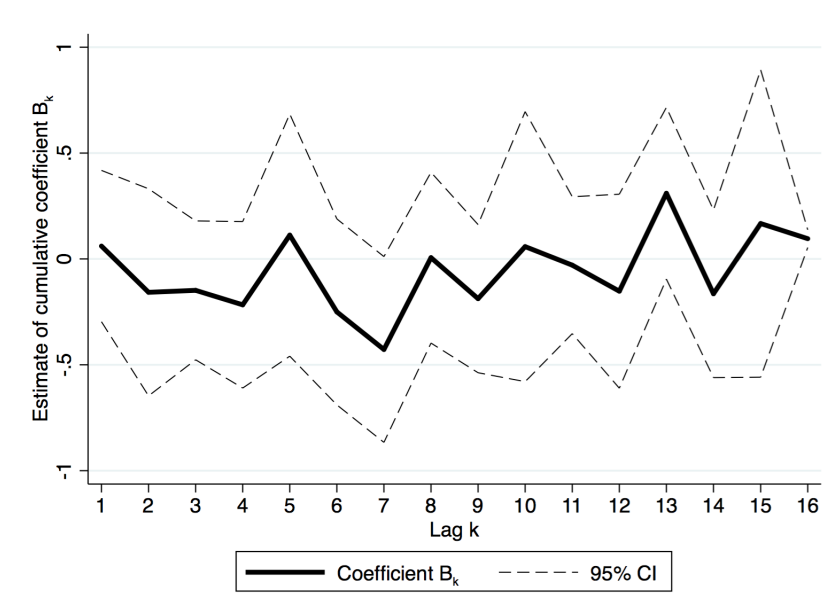
Notes: Plot of the lagged coefficients β_k from the specification 4.5. Note that β_k is printed in bold, which indicates that this is the estimate of the cumulative effect of the oil price lags on the outcome variables and is essentially the sum of the β_k up to lag k in specification 4.5 (see the discussion on the linear transformation of the lags in the previous section).

in the beginning and at the end of our sample. This should give us the first intuition regarding the change in the spending patterns and, essentially, represents the simplest form of the difference-in-difference approach. Results are reported in Table 4.1. One can see that the share of expenditures on food and on daily non food items decreased significantly during the oil boom. This is consistent with Engel's law which suggests that income elasticity of demand for food is between 0 and 1. On the other hand, expenditure shares on services and education seemed to have increased significantly during the resource boom, suggesting an income elasticity of above one. In particular, expenditure on education doubled during the resource boom in oil rich districts, increasing the share spent on education from 3% to 6%. Figure 4.3 plots our main variable of interest (binary variable, which equals 1 if a household reported the payment of tuition fees for graduate studies) over time. Thus, it looks as if the oil boom was indeed associated with growing educational spending on tuition fees in graduate education.

A more thorough analysis is based on specifications presented in the previous section. The results from estimating 4.5 using the logged real wage and the probability to be employed in the formal sector are presented in Figures 4.5 and 4.4, respectively. We plot the lagged coefficients β_k from the specification 4.5. Note, however, that β_k is printed in bold, which indicates that this is the estimate of the cumulative effect of oil price lags on the outcome variables and is essentially the sum of the β_k up to lag k in specification 4.5 (see the discussion on the linear transformation of the lags in the previous section). Thus, Figure 4.4 suggests that the cumulative effect of oil price fluctuations becomes significant after slightly more than a year. The coefficient suggests that a 10% increase in the oil price increases the probability of being employed as an educated worker in the formal sector relative to an uneducated worker by 1 percentage point after approximately 6 quarters. Figure 4.5 plots the estimate of the cumulative coefficients for the wage regression and shows that it is not affected by the resource boom. While it fits our model, this may be partly driven by the fact that the resource boom triggers a substantial immigration of educated labour force into resource-rich districts, thus diluting the possible wage response; this will be further discussed in the next section.

The results from estimating specification 4.6 are presented for alternative specifications

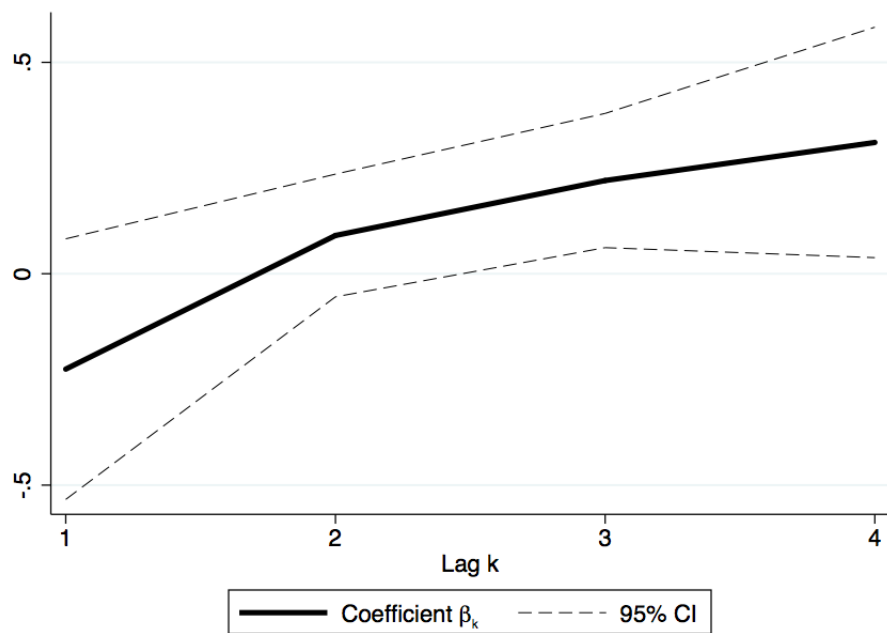
Figure 4.5: Real Quarterly Wage in the Formal Sector (logged)



Notes: Plot of the lagged coefficients β_k from the specification 4.5. Note that β_k is printed in bold, which indicates that this is the estimate of the cumulative effect of the oil price lags on the outcome variables and is essentially the sum of the β_k up to lag k in specification 4.5 (see the discussion on the linear transformation of the lags in the previous section).

in Table 4.2 in the Appendix. Most of the specifications will be discussed in the next section on the robustness of our results. Here, the main result (column two in Table 4.2) is plotted analogously to the results above in Figure 4.6 (again, using annual and not quarterly data). The results are unambiguous: an oil price boom leads to an increase in the probability of a household paying tuition fees with a two year lag. In particular, a 10% increase in the oil price increased the probability that households pay tuition fees by 3 percentage points after approximately 3 years. This should be interpreted as evidence that at least some members of the household have begun studies at a university or vocational training institution. The results are entirely consistent with the theoretical model we developed and, ultimately, demonstrate that the resource boom in Kazakhstan triggered both an increase in employment opportunities for educated worker and also in demand for education.

Figure 4.6: Probability that a household pays tuition fees.



Notes: Plot of the lagged coefficients β_k from the specification 4.6. Note that the β_k is bold, which captures the fact that this is the estimate of the cumulative effect of the oil price on the outcome variables and is essentially the sum of the β_k up to lag k in specification 4.6 (see the discussion on the linear transformation of the model in the previous section).

4.7 Discussion

4.7.1 Is our result driven by the Engel curve effect?

The results reported so far are definitively consistent with our original theory. However, we still have to rule out another equally important explanation: our results may simply represent the Engel curve effect. Since, oil-rich districts experienced a faster growth in income, one could expect the share of expenditure in superior goods (including education) to increase relative to other expenditure. In what follows we present some evidence that this is unlikely to drive the results.

Engel curves in resource-rich and resource poor districts: Assume that we look at two households in oil-rich and in oil-poor districts, who have exactly the same absolute level of expenditure. Then these households, given the Engel curve effect, should have the same share of educational expenditure. This does not mean that the aggregate expenditure for education in oil-rich and oil-poor districts are the same: the population of oil-rich districts may include a larger number of households with higher expenditure, spending relatively more on education. The match should persist throughout the resource boom: its only effect would then be growing expenditure of the population, translated into a larger number of high-expenditure individuals with a larger share of expenditure on education. Graphically, if we plot the Engel curves (i.e., the effect of income on the share of expenditure on tertiary education) for both groups, these graphs should coincide. On the other hand, if households in the oil rich districts have a higher incentive to invest in education, the Engel curve for educational expenditure should shift upwards after the beginning of the boom. This is theoretically captured in equation 4.4.

Figures 4.7 and 4.8 depict the shape of Engel curves for the resource-rich and resource-poor districts in 2001 (i.e. the beginning of the resource boom) and 2002-2005 (during the boom). The coefficient of logged household expenditure is estimated non-parametrically (kernel-weighted local polynomial smoothing) after partialing out household fixed effects, time fixed effects and household size. As shown in Figure 4.7, the Engel curves of both groups

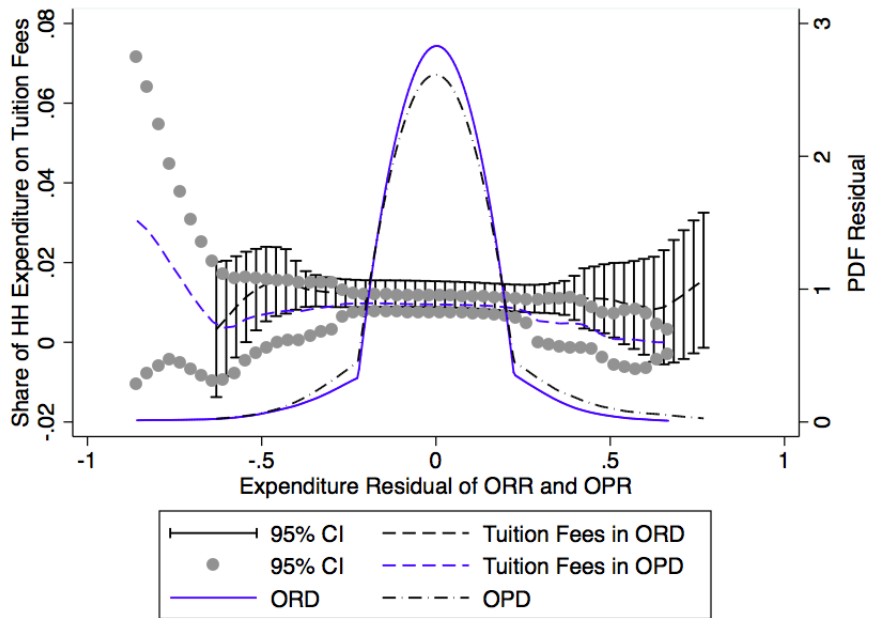
are almost identical in 2001. After 2001, however, the situation looks entirely different as presented in Figure 4.8. The Engel curve for the resource-rich districts is significantly above the Engel curve for the oil poor districts for the central part of the distribution. This suggests that, from the beginning of the boom households in the resource-rich districts spent more on tertiary education than comparable households in the resource-poor districts.

Income and expenditures: An indirect argument against the Engel curve interpretation of our results can be extracted from Toews (2014), who shows that in Kazakhstan in oil-rich districts the increase of reported income as a result of the resource-boom was substantially stronger than the increase of reported expenditure, which essentially remained the same. If that is the case, we should not expect the Engel curve effects to drive our results. Toews (2014) explains his finding by the shift of employment from the informal to the formal sector, which made people more honest in acknowledging their income in the survey. This is also a possible interpretation of our results: as mentioned, we define employment merely as 'formal employment'; the alternative is likely to be not only 'pure' unemployment, but also various forms of informal labour activity. However, educated workers typically receive a larger premium from formal employment relative to that received by uneducated workers (e.g. Günther and Launov (2012)). Hence, the increase of opportunities for formal employment (even if these mean a shift from the informal to the formal sector) should be an incentive strong enough to stimulate educational expenditures.

4.7.2 Is our result driven by increased supply of education?

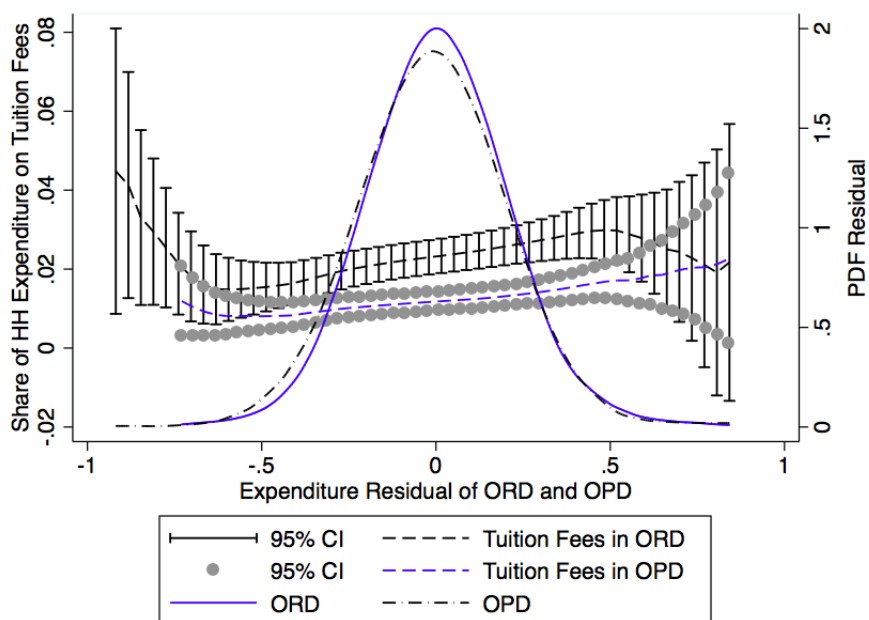
The growing educational expenditures may have been driven by a supply effect. This implies that the oil boom resulted in a substantial growth of the availability of tertiary education (e.g. new universities and technikums were established, more funding to the educational system was provided), and as a result people received greater opportunities to study. Stated otherwise, the supply effect still suggests that there is a positive link from resource booms to education. However, it is not driven by the demand of households, but by the increased provision of public goods by the government. In our model this would be equivalent to a decrease in Θ . Much of the tertiary education in Kazakhstan is still provided by the government, and

Figure 4.7: Engel Curve in 2001



Notes: Plot of the non-parametric estimate (kernel-weighted local polynomial smoothing) of the slope coefficient of logged household expenditure after partialing out time fixed effects, household fixed effects and household size. The distribution of logged household expenditure after partialing out time fixed effects, household fixed effects and household size is plotted in the background.

Figure 4.8: Engel Curve after 2001



Notes: Plot of the non-parametric estimate (kernel-weighted local polynomial smoothing) of the slope coefficient of logged household expenditure after partialing out time fixed effects, household fixed effects and household size. The distribution of logged household expenditure after partialing out time fixed effects, household fixed effects and household size is plotted in the background.

most decisions regarding the supply of tertiary education are made at the central or provincial levels. District authorities have negligible powers in the political hierarchy in Kazakhstan. More importantly, provinces are a typical unit of educational planning (World Bank 2005b). Thus, any supply effect is invariant for all households located in the same province in a particular period of time.

There are at least two ways to deal with the possible supply effect: quantity and quality of the tertiary education provided. In the first step we control for the number of universities and technikums in a particular province and year. The results of this specification are presented in the third column of Table 4.2 in the Appendix. Controlling for the number of tertiary institutions on the province level does not affect our estimated coefficients significantly. From the raw data presented in Table 4.6 in the Appendix, we know that the numbers of universities and technikums were quite similar and followed a similar development. The number of technikums in five provinces we focus on increased by 31.9%. In comparison, the number increased by 32.2% in the 'capital cities' (Almaty and Astana)¹³ and in the rest of the country (resource-poor provinces) by 30.1%. The number of universities has barely been affected across provinces.

Second, the supply effect was possibly driven by an improvement in the quality of education provided in different parts of Kazakhstan. Here we have to rely on province-level analysis; however, as mentioned, decisions regarding educational expenditure are made at province level. Since 2008, Kazakhstani government has been publishing a rating of provinces of Kazakhstan in terms of education. The rating is based on several proxies (access to education, staff, financial and material resources and outputs), but most precisely capture the quality of educational services supplied. The oil-rich provinces, however, do not score well in these ratings. Atyrau in 2008-2010 scored 10th, 9th and 9th respectively out of 16 provinces; Kyzylorda scored 14th, 14th and 11th; Mangistau scored 12th, 10th and 12th. The results

¹³Throughout the Soviet period, the capital of the Kazakh Soviet Republic was Almaty. It remained the capital of independent Kazakhstan until 1997, when the government was transferred to the new capital, Astana. However, Almaty remains an important business and education hub, and thus should be treated, similarly to Astana, as a special region of Kazakhstan. In particular, most important universities have been traditionally established in the capital cities.

are somewhat better for Aktobe (5th, 8th and 14th) and for Western Kazakhstan (8th, 5th and 6th), but even these provinces do not show impressive performances, as opposed to some oil-poor provinces, like Pavlodar, Kostanai or Karaganda (Ministry of Education of Kazakhstan 2011). One could of course hypothesize that the situation in the oil-rich districts was even worse before the start of the boom. However, this is unlikely because Kazakhstan inherited its educational system from the Soviet Union, and in the USSR the regional variation the educational quality (except for a very few leading centres) was very low due to the targeted efforts of the central government.

More generally, educational government expenditure on the province level implies that all supply changes should be captured by province-time fixed effects. The results of this specification are presented in the fourth column of Table 4.2 (see Appendix). If anything, the results suggest that the effect becomes stronger. Certainly, the private educational sector is not constrained by these institutional features in its development; but it is unlikely that private educational facilities are created unless there is a demand for their services, and therefore the supply effect cannot be driven by this group.¹⁴

4.7.3 Is our result driven by educated immigrants?

One could argue that more educated households moved into the oil rich districts during the boom. If educated households are more likely to send their children to universities, the proportion of households reporting the payment of tuition fees should increase naturally. Thus, despite controlling for household fixed effects in our main specification, migration is an important confounding factor in our analysis and must be discussed.

On the one hand, while generally the inter-provincial migration in Kazakhstan is not very high,¹⁵ it is likely to be much higher within provinces. On the other hand, large spatial dis-

¹⁴Endowment universities and other forms of private schools operating through generous funding and thus not bound to short-term profit maximization are almost unheard of in Kazakhstan.

¹⁵Still, according to the Census data, in 1999-2009 internal migration in Kazakhstan increased almost fivefold; internal immigration to Kyzylorda and Atyrau provinces increased over-proportionally by 12.1 and 9.8 times, respectively (Sange Research Center 2009).

tances across the regions of Kazakhstan¹⁶ ensure that, for example, inhabitants of resource poor districts cannot accept a job in an oil or oil-related sector without migrating to the districts where the oil is extracted; therefore migration costs should have an effect on employment decisions. Unfortunately, we do not have information on the migration on district level. To account for migration, we reestimate our main specification by focusing on those households which had remained part of the survey since 2001. The results are presented in column 5 of Table 4.2 in the Appendix. The results suggest a larger and still significant cumulative effect in comparison to our main specification. This is quite intuitive: households residing in the oil rich districts since the beginning of the boom did not have to consider migration costs and thus were more likely to invest in education. Note, however, that we obtain a significant and negative contemporaneous effect of the resource boom on the probability to invest in education. The intuition behind this result is likely to be the following. Following the increase in the oil price the first and most straightforward result is an increasing demand for the labour force, which decreases the willingness of *local* population to study given the immediate benefits to realize on the labour market (for the population in other districts the effect is absent - to benefit from resource boom, they need to migrate first, and it does not happen immediately). In the longer run, however, both local and non-local population start investing in education, to benefit from the opportunities provided by the resource boom to the educated labour force.

4.7.4 Is our result driven by peer effects and educational signaling?

The discussion so far has been based on the assumption that the growing oil sector is more likely to employ *any* individual with a tertiary education (university or vocational training degree). This point of view is entirely applicable if we look at education merely through the 'educational signaling' perspective, i.e. education does not result in increase of human capital, but rather makes it possible for the employer to separate the high-quality from the low-quality applicants. Otherwise, however, oil companies could be more likely to employ educated workers, but only if their education profile fits the demands of an oil company. This profile does not necessarily include only oil-related degrees: oil companies may be also in-

¹⁶The territory we are looking at is 4 times the size of UK or about 10% of the US.

terested in hiring general engineering or IT specialists or accounting and business experts. Unfortunately, Kazakhstani oil companies do not disclose detailed information on the educational background of their employees, and our data do not contain information on what the members of the households in our sample actually study. Still, in this section we attempt to collect some information on how the study profiles fit the possible demand of the oil industry to refine our conclusions.

To start with, as mentioned, in our data we define as 'education' two types of degrees: university education (which was originally based on the Soviet-style 5 to 6 years 'specialist' programs, and is currently organized in the BSc / MSc programs) and the vocational training (in 'technikums'). However, in our sample only 10-15% of the population has a university degree; 30-40%, in contrast, have a technikum degree. Thus, it is most probable that our results are driven by those receiving education in technikums. Moreover, it is quite common in Kazakhstan to move to the capitals to study at university. Figures 4.10 and 4.11 in the Appendix report the change in the number of students at the universities and technikums in three areas of Kazakhstan: five oil-rich provinces, two 'capital cities' and the remainder of the country. For technikums, the results are straightforward: throughout the last decade, the number of technikum students in the oil-rich provinces was growing faster than in the rest of the country. Occasionally, it grew slower than in Astana and Almaty; however, one has to recognize that the main educational centres of the country are located in these cities, which therefore train many more students than are trained in any other part of the country (even the oil and gas industry heavily relies on educational facilities in the capital cities). At the same time, in terms of the number of students studying at universities, oil-rich provinces underperform as opposed to oil-poor provinces. Certainly, it is possible that students from oil-rich provinces study in other parts of the country as well (most likely in the capital cities), but, again, this confirms that the results we observe are probably driven primarily by technikums.¹⁷

¹⁷Interestingly, in Russia, which is probably the closest case in terms of comparison to Kazakhstan, we obtain a very similar outcome: in Tyumen province, which is the main oil extracting region of the country (and is also located in a rather hostile environment) the number of technikums increased between 2000 and 2003 by 30.1%, while the average growth rate in the Russian Federation excluding Moscow and St. Petersburg was 8.7%; in Kazakhstan the respective indicators were 23% for oil-rich and 20.8% for oil-poor provinces excluding capital cities.

The basic design of the Kazakhstani system of technikums is, as mentioned, inherited from the Soviet Union. These institutions concentrate on training students to become qualified workers and craftsmen. However, there have been certain changes in terms of enrolment rate, in specialization and in the financing mode, all of which we have to take into account. Most importantly, the structure of specializations undergone by the technikum students greatly changed during the post-Soviet era. UNDP (2004) claims that, for Kazakhstan, 60% of technikum students study the humanities, business or teaching. There is no detailed data on the district level, but some information can be obtained if we look at the number and specializations of technikums in various parts of Kazakhstan. As of 2012, five oil-rich provinces contained 96 public and 58 private technikums. Of these public technikums, only 7 specialized in the oil, gas and energy professions; and only one in construction. Fifty technikums did not list any particular specialization, and some also are known to offer more oil-related specializations (among other specializations). Twelve technikums, however, offered education in the humanities, the arts, business and teaching. In the private technikum sector, there was only one specialized oil and gas technikum (there were 2 construction technikums as well) versus 18 general-purpose and 29 humanities, law, arts and business technikums. Of course, the number of technikums may not be representative for the number of students, but it is still unlikely that training for the oil and gas professions dominates the educational structure of the students.

Thus, in order to explain the actual growth of demand for education in Kazakhstan, we may need to adjust the original model. There are three possible explanations for what we find. First, as mentioned, is the educational signal argument: not the content of education, but the mere fact of having an education (and thus distinguishing oneself from other labour market participants) matters. It is important to point out that in many transition countries the link between actual specialization and the occupation became diluted during the post-Soviet transition, when the labour markets underwent fundamental change: we may be capturing the consequences of this effect. Second, the behavior of students may be driven by peer effects. For example, if 'the best' students go to oil and gas related technikums and can benefit greatly from a higher income, other students may go to technikums merely imitating the be-

behaviour of the best or in an attempt to keep up their social networks, even if their employment opportunities actually do not go up. There is evidence of this type of behaviour in developing countries as reported e.g. by Bobonis and Finan (2009), Kremer and Holla (2009). Third, Nguyen (2008), Jensen (2010a) and Emerson and McGough (2011) point out that the crucial feature in demand for education is not the actual education premium itself, but rather the expected returns to education, as assumed in our model. As discussed above, in many developing countries, this rather acts as a constraint on the demand for education. However, one could argue that, during an oil boom, expectations become overly optimistic, if fueled by the observation of the actually growing demand of the oil industry for any kind of qualified labour.

We should stress that, if the second and third explanations are true, the demand for educated labour from the oil companies is still great enough to drive the observed increasing employment probability for educated labour in oil-rich districts. Nevertheless, we must point out that the actual mechanism driving the results of the study may be more complex than the stylized model we reported.

4.8 Conclusion

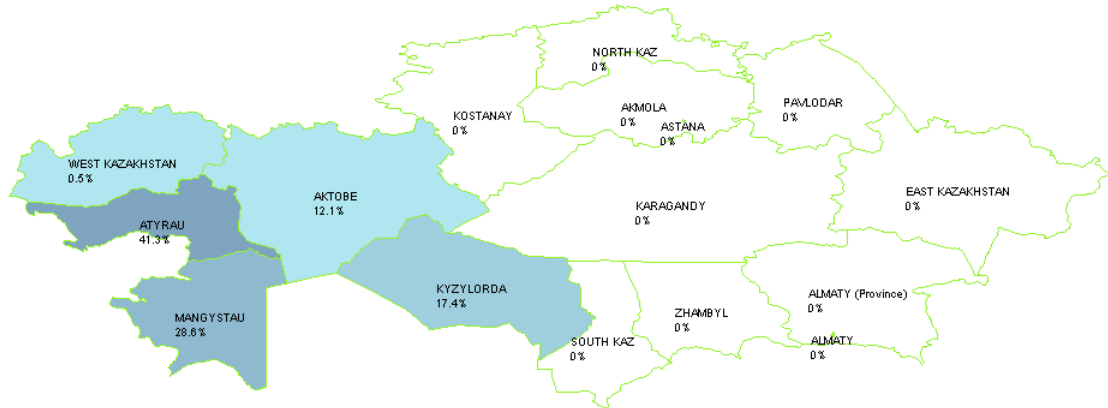
The paper intends to investigate whether increasing job market opportunities for educated labour may trigger private demand for education. Using a unique micro-level panel dataset from Kazakhstan, and applying the oil boom this country experienced in 2001-2005 as the identification strategy, we indeed confirm that the private demand for education went up. Our results are robust to various estimation techniques and controls. We are also able to show that the effect we see is not driven merely by the fact that education is a superior good and thus demand for education goes up simply because earnings go up. Our subsequent discussion suggests, however, that the increased demand for tertiary education was not driven by students studying engineering or construction only. Instead, our results suggest that peer effects and overly optimistic assessment of perceived returns to education might have played an important role in increasing demand for education.

We find that following the increased probability of being employed as an educated worker, the probability that households pay for tertiary education increases. There are two major implications of our results. First, more generally, we confirm that lack of employment opportunities constitutes a major constraint in demand for education in developing countries. Removing this constraint (or even providing a credible signal that it is going to be removed) may have a strong positive effect on demand for education. However, we acknowledge that in Kazakhstan, first, costs of education are not prohibitive due to the remaining educational infrastructure of the Soviet period and, second, the population already possesses a good basic education level. These assumptions do not always hold in developing countries and may constrain the effectiveness of the mechanism we have described in our paper.

Second, our paper also contributes to the general discussion of how resource boom can affect human capital accumulation. Basically, we claim that private households can at least partly compensate for the well-known inefficiencies of the government in terms of dealing with resource revenues. In our case, inefficiencies which are related to education expenditures. However, this holds only under certain conditions, most prominently strong linkages of the oil sector to the local labour market (i.e. resource boom triggers increase of labour opportunities) and relatively high educated labour intensity of the oil sector (due to technological and environmental challenges). For many developing countries these conditions may not be fulfilled, as the role of the local labour force in the oil extraction is very low. Furthermore, in our case there was very limited direct redistribution of oil gains for welfare purposes (Kazakhstan's social security system is very weak); if this were greater, it could possibly discourage educational investments as well. Thus, our results can be seen as a set of conditions on how the private demand for education can be generated in the case of resource boom; whether these conditions hold depends on political economy considerations.

4.9 Appendix

Figure 4.9: Share of National Oil Production



Source: SARK (2011).

Table 4.2: Dependent variable: Binary Variable indicating expenditure on tuition fees for tertiary education

	FE1	FE2	FE3	FE4	FE5
	(1)	(2)	(3)	(4)	(5)
β_1	-0.226 (0.157)	-0.227 (0.157)	-0.246 (0.165)	-0.348 (0.29)	-0.520 (0.234)**
$\beta_1 + \beta_2$	0.09 (0.074)	0.092 (0.073)	0.094 (0.074)	0.262 (0.117)**	0.135 (0.093)
$\beta_1 + \beta_2 + \beta_3$	0.22 (0.081)***	0.218 (0.081)***	0.206 (0.081)**	0.306 (0.146)**	0.233 (0.133)*
$\beta_1 + \beta_2 + \beta_3 + \beta_4$	0.311 (0.139)**	0.306 (0.139)**	0.304 (0.141)**	0.616 (0.279)**	0.53 (0.248)**
Controls	N	Y	Y	Y	Y
Tertiary Institutions	N	N	Y	N	N
Province-Year FE	N	N	N	Y	N
Reduced Sample	N	N	N	N	Y
Obs.	6521	6521	6521	6521	2308

Notes: Standard errors adjusted for clustering at the district level are in parentheses : *** p<0.01, ** p<0.05, * p<0.1. Household fixed effects, time fixed effects, household size and household size squared are included in every regression. The full set of household head characteristics includes education and occupation. In column 3 we control for the number of tertiary institutions on the province level. In column 4 we control for province year fixed effects. In column 5 we keep only households which have been part of the sample since the beginning of the boom (2001).

Table 4.3: Accumulative Production of main oil and gas consortia between 2001-2005 (Oil and Gas in 1000 tonnes), ownership and location.

Name	Major Owner	Cum. Prod.	Share (in %)	Province	District	Webpage
Kazakhstan	-	265000	100%	Aktobe, Atyrau, Kyzylorda, Mangystau, West-Kazakhstan	-	-
Tengiz-Chevroil	Chevron, Exxon-Mobil, KazMunayGas, LukArco	65500	25%	Atyrau, Mangystau	Zhylyoyskiy (Atyrau), Beyneuskiy (Mangystau)	www.tengizchevroil.com
EmbaMunay Gas	KazMunay-Gas	13000	5%	Atyrau	Kzylkoginskiy, Isatayiskiy, Makhambetskiy, Makatskiy, Zhylyoyskiy, Atyrau (city)	www.kmg.kz
UzenMunay Gas	KazMunay-Gas	27000	10%	Mangystau	Manghystauskiy, Karakhiyanskiy, Zhanozen (city)	www.kmg.kz
Mangystau-MunayGas	Central Asia Petroleum Ltd. (until 2009)	25000	9%	Mangystau	Manghystauskiy, Karakhiyanskiy, Tupkaraganskiy, Zhanozen (city), Aktau (city)	www.mmg.kz
Karazhanbasmunai	Canada's Nations Energy Ltd. (until 2006)	10000	4%	Mangystau	Tupkaraganskiy, Manghystauskiy	www.kbm.kz
Karashaganak	BP Group, ENI	36000	13%	West-Kazakhstan	Burlinski	www.kpo.kz

Source: Munayshy Public Foundation (2005).

Table 4.4: Accumulative Production of main oil and gas consortia between 2001-2005 (Oil and Gas in 1000 tonnes), ownership and location. (Cont. Table 4.3)

Name	Major Owner	Cum. Prod.	Share (in %)	Province	District	Webpage
Actobe -munaigas	CNPC	24000	9%	Actobe	Temirskiy, galzharskiy, Baiganinskiy	Mu- www.cnpc.ch
Kazakhoil Actobe	KazMunay Gas, LukOil	3500	1%	Actobe	Temirskiy, galzharskiy, Baiganinskiy	Mu- www.koa.kz
Petro - Kazakhstan	Petro - Kazakhstan (until 2005)	23000	9%	Kyzylorda	Syrdarinskiy	www.petro -kazakhstan.kz
Petroleum -Turgai	LukOil, Hurricane Kumkol Munai	13000	5%	Kyzylorda	Syrdarinskiy	www.turgai.kz
KasGer -munai	JSC Yuzh-neftegas, Feba Oil AG, Erbdol Erdgras Gommern	8500	3%	Kyzylorda	Syrdarinskiy	www.kazger -munay.kz

Source: Munayshy Public Foundation (2005).

Table 4.5: Description, mean and standard deviation of the main variables on the individual and the household level.

Variables	Description	Mean (ORD)	SD (OPD)	Mean (OPD)	SD (OPD)
Sample Size	Number of Household observations Cross-section *Time periods (N*T)	3069	-	6036	-
Yearly real household expenditure (logged).	Aggregated yearly household expenditure in Tenge. Provincial inflation from the SARK (2011) is used to discount the values to the end of 2000.	12.62	0.56	12.34	0.52
Education Share	Share of total expenditure spent on tuition fees	0.02	0.05	0.014	0.04
Household Size	Total number of people in a household.	4.7	2.3	4	2
Sample Size	Number of Individual observations Cross-section *Time periods (NT)	33084	-	53520	-
Higher Education	Proportion of individuals who received a higher education	0.11	0.32	0.16	0.36
Vocational Training	Proportion of individuals who received vocational training	0.37	0.48	0.32	0.47
No Tertiary Education	Proportion of individuals who received up to complete secondary education	0.52	0.5	0.52	0.5
Employment	Share of those who reported to be employed in the formal sector (public sector or firms in the private sector)	0.45	0.5	0.43	0.49
Quarterly real wage (logged)	Quarterly real wage of those employed in the formal sector. Provincial inflation from the SARK (2011) is used to discount the wage to the end of 2000	10.5	0.82	10.1	0.74

Figure 4.10: Growth rates of the number of students enrolled in technikums in Kazakhstan, based on official statistics

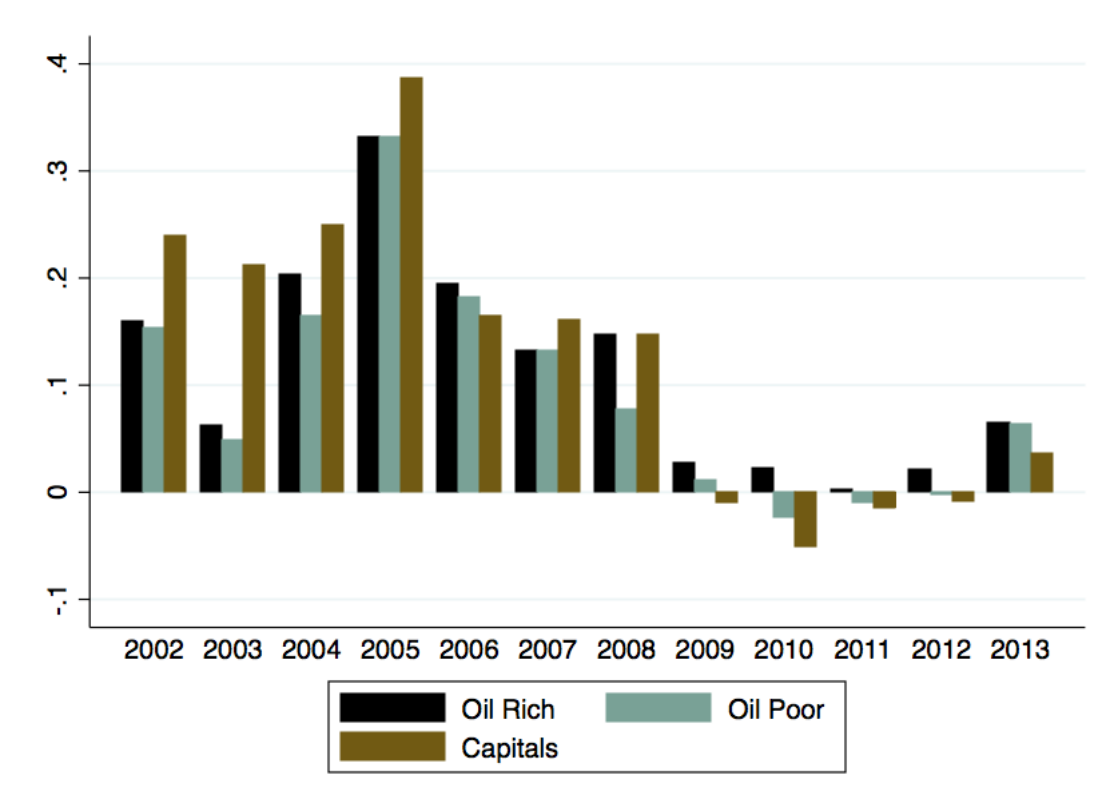
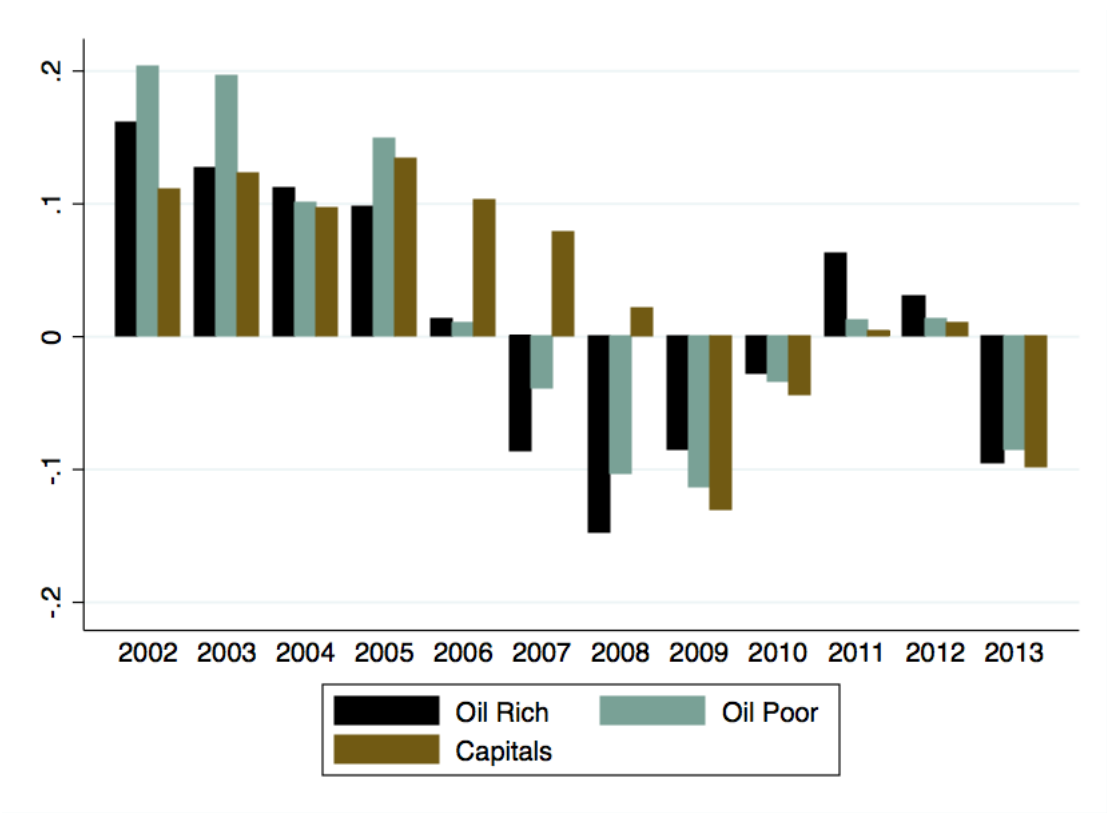


Table 4.6: Number of Technikums/ Universities in oil-rich provinces of Kazakhstan

Years	Number of Technikums					Number of Universities				
	2001	2002	2003	2004	2005	2001	2002	2003	2004	2005
Aktobe	13	13	13	15	16	6	6	6	7	7
Atyrau	9	8	10	11	11	3	3	3	3	3
Kyzylorda	11	14	15	18	18	8	7	6	6	6
Mangystau	11	11	12	13	14	4	4	5	4	3
West Kaz.	9	9	10	10	10	5	5	6	6	7

Figure 4.11: Growth rates of the number of students enrolled in universities or equivalent educational institutions in Kazakhstan, based on official statistics



Chapter 5

Mine closures and female employment: Evidence from the UK

5.1 Introduction

The emergence of the coal industry in the 18th century and its impact on UK's industrialization has been the topic of a long-lasting debate (Fernihough and O'Rourke 2014). At the same time, barely anything is known about the long-term consequences of its abrupt end. Our goal is to contribute to filling this gap.

We exploit the dramatic run-down of the coal industry in the UK since 1981 to study labour markets which have been affected by the coal bust. The specific feature of a negative labour demand shock originating in the mining industry is the predominance of males being laid off. This is because mining is traditionally a male-dominated activity.

An important contribution of our paper is the construction of a new data set containing information on the location of coal mines, the number of miners active within individual coal mines, and the year of closure from 1981 onwards. We spatially link this data set to the UK census for the period 1981-2011. Thus our data covers the total population of people and miners in England and Wales and consists of 339 districts, 211 mines and 4 periods (the UK Census is collected every 10 years). Out of a total of 211 mines in 1981 almost all were shut

down by 2011. Consequently, the number of active miners dropped from approximately 190 thousand in 1981 to less than 3 thousand in 2011. We use this locally dispersed labour demand shock as our treatment affecting labour markets in districts which had at least one mine since 1981, our treatment group. Our control group consists of districts without a coal mine in 1981.

We find that districts affected by the coal bust experience a rapid and permanent relative decrease in the population. We do not observe an increase in unemployment in the effected districts. This is because individuals are willing to migrate to equalize employment opportunities, exit the labour force or find a job in other industries. Most interestingly, on the aggregated level (males and females) we find evidence for the first two channels but do not find any significant relationship with coal-mine closures and employment in manufacturing or services. Disaggregating the data by gender, however, suggests that males laid off during mine closures substituted female workers in other industries.

The results are important because it is a well-established fact that females' position in society crucially depends on their labour-force participation (Ross 2012). Moreover, we know that, *ceteris paribus*, resource rich economies usually reach a point at which the extraction of their exhaustible resource stops being economically feasible. If the extraction of a natural resource is dominated by males, such a bust will most certainly result in a gender-biased negative labour-demand shock which might result in a crowding out of female workers in other industries. It is important to take this mechanism into account when planing the transition from being a resource rich economy because gender equity is associated with large economic and social gains (Elborgh-Woytek et al. 2013).

Theoretically, this paper is most closely linked to Moretti (2011) who studies shocks to labour markets with heterogeneous labour. Empirically, the paper is most closely related to the currently developing literature on local impacts of natural booms and busts (Black et al. 2005a; Ross 2012; Aragón and Rud 2013; Allcott and Keniston 2013). Ross (2012) is maybe the first to link the role of females in the society to the abundance of natural resources. How-

ever, he suggests that female participation drops during a boom rather than a bust because of the Dutch Disease and generous government transfers which decrease the incentive to work. Finally, the paper is also related to Beatty and Fothergill (1996) and Beatty et al. (2007) who were the first to document that many miners laid-off during the coal bust in the UK exited the labour force but did not migrate.

Our work is novel in three respects. We are the first to construct a spatial data set capturing all active coal mines in the UK since 1981. Second, we use the data in a quasi experimental framework to study labour markets in districts affected by coal mine closures. Third, we provide some evidence that the closure of coal mines since 1981 is associated with a persistent reduction in female labour force participation.

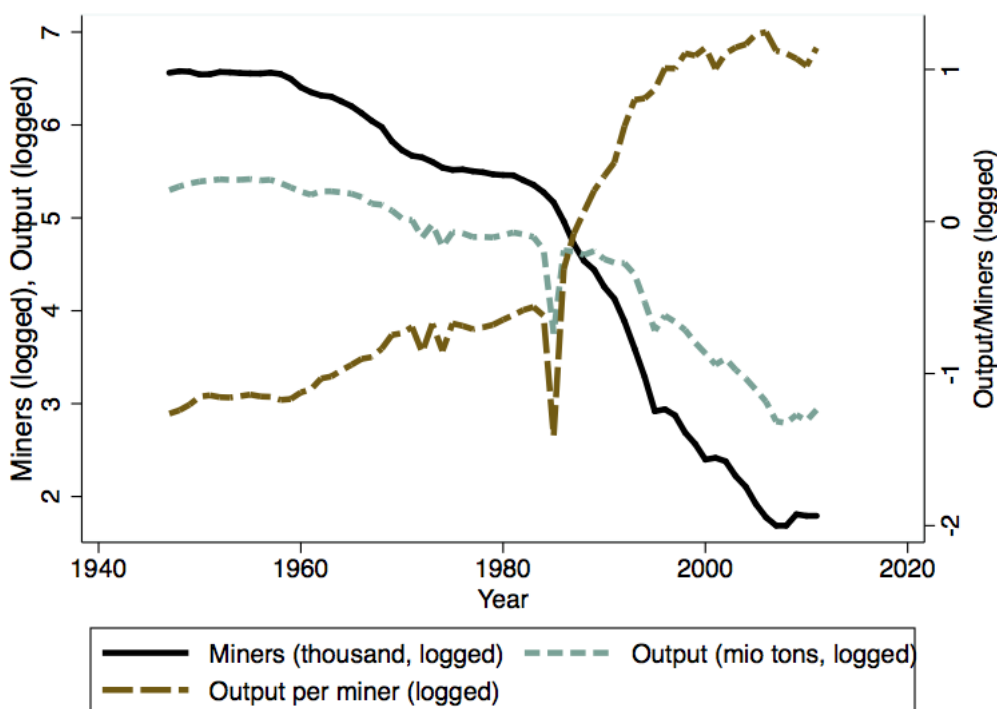
The paper is structured as follows. In the next section we present a brief history of the coal industry in the UK. In the third section we derive a simple theoretical framework. In the fourth section the data employed in the empirical analysis is presented. In the fifth section we discuss our identifications and estimation strategy. Results are discussed in the sixth section before we conclude.

5.2 Empirical Case: Coal Industry in the UK

Historically, the extraction of coal has played an extraordinary role in the UK's history (Fernihough and O'Rourke 2014). It played a key role in the process of the UK's industrialization and the subsequent economic growth. It has also been important in determining the allocation of economic activity within the UK since the 18th century.

Employment in coal mining reached its peak around 1920 with over 1.2 million active miners (Cook and Stevenson 2001). Since then, employment and output has been steadily decreasing. In 1947 the coal industry was nationalized under the National Coal Board (NCB). The NCB governed the coal industry until 1994 when the remaining pits were either closed or

Figure 5.1: Output, Employment and Productivity



privatized. After WWII, industrial relations in the coal industry were established by the NCB and its counterpart the National Union of Mineworkers (NUM), created in 1945.¹ As opposed to the NCB, the NUM is still active, even though it lost much of its power in the miners' strike of 1984-85.

There is little doubt about market forces playing a key role in the decreasing demand for domestically mined coal and the collapse of the coal industry since 1945, shown in Figure 5.1 (Parker and Surrey 1993). In 1950, coal mined domestically accounted for 90% of the total of the UK's primary energy consumption (Surrey 1992). Throughout the 1950s, the total number of miners remained around 700000, and annual output of coal was above 200 million tons. Despite technical advances, the industry entered a recession from the mid 1960s onwards. The cease in demand for coal was mainly driven by an increased availability of substitutes, most prominently crude oil and nuclear power, and increased environmental concerns (Sur-

¹The roots of NUM go back as far as the end of the 19th century (NUM 2014).

rey 1992; NUM 2014). Consequently, coal mined domestically accounted for only two-thirds of total primary energy consumption in the early 1970s, and major closure programs forced annual output to fall to 130 million tons and employment to 270000.

The recession ceased abruptly following the oil price shock in 1974. In combination with major delays in the supply of nuclear power, the electricity supply industry demanded all the coal the NCB could produce (Surrey 1992). In 1974 the NCB and the mining unions agreed on a plan of long-term investment to ensure a total of 120 million tons of deep-mined capacity by 1985. Investment continued to rise rapidly from 1975 to 1979, when the second oil shock drove oil prices up even further and indicated a higher future demand for coal.

Nevertheless, demand for coal and eventually its output continued to decrease throughout the 1980s such that by the beginning of the 1990s domestic coal accounted for less than 20% of total primary energy consumption. At the beginning of the 1980s the decrease in demand was driven by the economic recession and the structural decline of heavy industries (Surrey 1992). In 1984 Margaret Thatcher famously declared war on the NUM by announcing: "We had to fight the enemy without in the Falklands. We always have to be aware of the enemy within, which is much more difficult to fight and more dangerous to liberty" (Thatcher 1993). Following the announcement of the closure of 20 pits, and a leakage of secret information about the planned closure of over 70 additional pits, the great strike began in March 1984 (Higham 2014; NUM 2014). The strike ended a year later in 1985 following a NUM vote to return to work (NUM 2014). The government's victory in the strike meant that the NUM's power was significantly diminished. Consequently, the initial orientation towards output maximization was replaced by the government's goal to maximize profits, the closure of unprofitable pits and huge lay-offs (Glyn 1988; Surrey 1992). Additionally, at the beginning of the 1990s the privatization of power plants and the subsequently increased competition from imported coal and natural gas ("dash for gas") have been major drivers in the decreasing demand, and eventually triggered the privatization of the mining industry in 1994 (Glyn and Machin 1997).

Thus, decreased demand for domestically produced coal played an important role in the

huge layoffs of miners. But several studies also suggest that technical change, the reallocation of workers in existing pits and a change in industrial relations benefiting the profitability aspect of coal extraction has increased productivity since 1985 (Glyn 1988; Surrey 1992; Glyn and Machin 1997). To a certain extent, increased productivity inevitably decreases employment. As a result it is difficult to assess how much of the manpower reduction since 1985 can be attributed to a decreasing demand for coal, the spread of technology² or a less constrained management style of the NCB following the change in industrial relations (Richardson and Wood 1989; Glyn and Machin 1997). But, due to some combination of these, the number of miners decreased from approximately 190000 in 1981 to less than 3000 miners in 2011. By 2011, domestic coal production was contributing less than 10% to total primary energy consumption and the closure of the next pit had already been announced.

5.3 Theoretical Framework

5.3.1 Baseline Model

This section presents a simple framework to analyze the impact of mine closures on local labour markets. The discussion is based on Moretti (2011), Greenstone et al. (2010) and Cordon and Neary (1982). We treat mine closures as a negative shock to the demand of local labour, and focus on their effect on the allocation of workers between industries and regions. This is a reasonable starting point, given the massive layoffs associated with mine closures in the UK case.

Consider an economy composed of several locations l . In every location there are two production sectors: mining and manufacturing which are denoted a and b respectively. All firms produce internationally traded goods with prices normalized to 1. Labour is the only variable factor of production and there is only one type of labour that can be used in both industries.

²Inter alia driven by the investment programs in the previous 10 years.

We assume that each worker provides one unit of labour, so that the population size is equal to labour supply, N . Workers are mobile and the indirect utility of worker i depends on wages in location l and an idiosyncratic preference over location: $U_{il} = w_l + e_{il}$. For simplicity we assume that there is no housing and that there are no amenities.³ As suggested by Moretti (2011), being born in location l and having family and friends make the location more attractive for worker i . Assume that there are only two locations, l_1 and l_2 , and that the relative preference of individual i over locations is $e_{il_1} - e_{il_2} \sim U[-\theta, \theta]$. The size of θ indicates the importance of preferences and gives rise to an upward sloping labour supply curve $w_{l_2} = w_{l_1} + \theta \frac{N_{l_2} - N_{l_1}}{N}$. In equilibrium the marginal worker is indifferent between locations. A larger θ indicates a lower willingness to migrate and thus a steeper labour supply curve. If $\theta = 0$ the labour supply curve is infinitely elastic.

In a given location, labour demand in each sector is $L_s(w, A_s)$ where $s \in \{a, b\}$. Labour demand depends on the local wage w and a sector-specific labour demand shifter A_s . Note that A captures firms' productivities but also other factors that may affect labour demand such as costs of production. Equilibrium in labour markets requires that $L_a(w, A_a) + L_b(w, A_b) = N(w)$.

What would be the effect of a demand shock in the mining industry, such as mine closures? In this framework, such a negative shock implies a reduction in A_a . Taking the total derivative we get

$$\frac{dw}{dA_a} = \frac{-\frac{\partial L_a}{\partial A_a}}{-\frac{\partial N}{\partial w} + \frac{\partial L_a}{\partial w} + \frac{\partial L_b}{\partial w}} > 0 \quad (5.1)$$

Equation 5.1 suggests that the closure of coal mines, a reduction in A_a , reduces local wages. The upward sloping labour supply implies an out-migration of workers such that local population decreases. There is a re-allocation of labour from mining to manufacturing. Due to the reduction in wages L_b increases, which captures the reverse Dutch Disease (Cordon and Neary 1982). The net effect on L_a is negative to account for the lower population size.

³Assuming no housing is equivalent to assuming a perfectly elastic housing supply. Relaxing this assumption does not change the qualitative predictions. As long as housing supply is not perfectly inelastic, the effect of demand shocks on population size is partially offset by an increase in housing costs. Assuming no amenities simply introduces an additional wedge between wages across locations and does not affect the qualitative predictions.

In this very simple scenario, mine closures unambiguously leads to an increase in the number of manufacturing workers and a reduction of the local population. This prediction changes significantly if we allow for agglomeration spillovers in the manufacturing sector.

5.3.2 Agglomerations

As discussed in Greenstone et al. (2010), agglomeration spillovers can affect firms' productivity and costs in several ways. First, the concentration of economic activity creates larger labour markets. Larger markets may facilitate job search and thus lead to better worker-firm matches. Second, concentration of economic activity can reduce transportation costs: in particular, transportation costs of local services and intermediate goods. Third, denser areas may facilitate knowledge spillovers. Sharing of knowledge and skills through formal and informal interaction may create positive production externalities across workers. Fourth, spatial concentration may be explained by natural advantages. In our case, the mining industry is located in regions with the most accessible coal fields. In either case, existence of agglomeration spillovers would imply that activity in the manufacturing sector is a function of local economic activity.

Following Moretti (2011), we assume that productivity in manufacturing depends on population size, $A_b = A_b(N)$. For simplicity, we assume that there are no agglomeration economies in mining. With this assumption, the equilibrium condition becomes $L_a(w, A_a) + L_b(w, A_b(N(w))) = N(w)$, where we make explicit that A_b depends on population size. Note that the equilibrium wage w^* is a function of A_a . With agglomeration spillovers, the effect of a mining shock A_a on manufacturing labour is unclear. In particular, we obtain that

$$\frac{dL_b}{dA_a} = \frac{dw^*}{dA_a} \left[\frac{\partial L_b}{\partial w} + \frac{\partial L_b}{\partial A_b} \sigma \right], \quad (5.2)$$

where $\sigma \equiv \frac{\partial A_b}{\partial N} \frac{\partial N}{\partial w}$ is a measure of the magnitude of agglomeration spillovers. Equation 5.2 highlights two channels through which mine closures (reduction in A_a) can affect manufacturing employment. On the one hand, the reduction of wages increases manufacturing em-

ployment. This is the reverse Dutch Disease effect. On the other hand, the reduction in population, and agglomeration economies, has a negative impact on A_b which reduces manufacturing firms' demand for labour. The net effect depends on the magnitude of agglomeration spillovers, σ . If σ is relatively small, then the effect of mine closures on L_b would be qualitatively similar to the case without agglomeration spillovers. However, if σ is sufficiently large then $\frac{dL_b}{dA_a} > 0$, and thus mine closures can *reduce* manufacturing employment.

5.3.3 Non-Tradable Goods

What is the effect of mine closures on non-tradable sectors, such as retail, catering and construction? These industries depend on local demand and are thus affected by the change in local aggregated income. Mine closures can reduce local income by reducing both population size and, if labour supply is upward-sloping, local wages. If non-tradable goods are normal goods the reduction in local aggregated income implies a reduction in non-tradable goods' production and employment.

The decline in wages, due to mine closures, can partially offset this decline. Mining workers can switch to non-tradable activities at lower wages. This general equilibrium effect, however, does not fully undo the reduction in aggregated demand for non-tradable goods (Moretti 2010).

5.3.4 Labour participation and unemployment

We have assumed that workers supply a fixed amount of labour. In reality, workers adjust the amount of labour in response to wages. In this case, the aggregated labour supply would be more elastic. In response to changes in wages, labour supply would change both due to migration of workers and due to changes in individual labour supply. We focus on the extensive margin. At the extensive margin, i.e. whether to work or not, we expect mine closures, and subsequent reduction in wages, to have a negative impact on labour participation.

The effect on long-run unemployment is less clear.⁴ Unemployment could increase for sev-

⁴In the short-run, we could expect an increase in unemployment. This is the standard result of negative demand

eral reasons. First, mine closures could increase structural unemployment. This could happen if mining workers have industry-specific skills that are difficult to transfer to other industries. Second, smaller labour markets may increase matching frictions of workers and firms. This is related to the loss of agglomeration economies associated with thick labour markets. However, other factors may actually *reduce* unemployment. For instance, layoffs and lower wages may discourage some workers to search for a job and exit the active labour force (e.g. early retirement). Even though they are not working, individuals who have exited the active labour force are not considered unemployed. Finally, the unemployment rate might remain the same if individuals migrate to equalize employment opportunities (Blanchard and Katz 1992).

5.3.5 Heterogeneous labour

Up to now we have considered only the case of identical workers. Here we allow for the existence of 2 types of worker: male and female. In particular, we assume an imperfect substitutability of males and females in the production process.

The mining sector is traditionally dominated by male workers. Thus, the closure of coal mines can be thought of as a negative labour demand shock to male workers only. Female workers could still be affected if males and females are imperfect substitutes in the production process. Male workers looking for alternative employment in manufacturing might crowd out females due to an increased supply and, thus, the relatively lower cost of male labour.

5.3.6 Empirical Predictions

This simple analytical framework provides several empirical predictions that we take to the data. In particular, it suggests that mine closures will:

1. reduce population size;

shocks in the presence of labour market frictions (Hall 1999).

2. reduce manufacturing employment if agglomeration spillovers are important. If they are not, then manufacturing employment increases;
3. reduce employment in non-tradable industries, such as retail, catering, and construction;
4. reduce labour participation, but have an ambiguous effect on unemployment;
5. reduce female employment if males and females are imperfect substitutes.

5.4 Data

For our main analysis we use two sources of data: a self-constructed data set on coal mines in the UK since 1981 and 4 periods of the UK Census (1981, 1991, 2001 and 2011). In what follows we discuss the structure of both data sets and the assumptions necessary to make data comparable across space and time.

An important contribution of this paper is the construction of a new data set containing information on location, the number of miners, and the year of closure of all active mines since 1981 in England and Wales. We include only mines which have been active as of the 1st of January 1981.⁵ The information on the location of pits is taken from the Guide to the Coal Field. The information on the number of colliers working within a pit comes from two different sources. For pits still active in 1984 the information is taken from Glyn (1988). Information on the number of colliers for pits which closed between 1981 and 1983 is taken from the Guide to the Coal Field. The year of closure of individual mines is taken from the Northern Mine Research Society and can be found on their webpage. The complete data set consists of 211 mines in 1981, most of which were closed within 15 years by the mid 1990s.

From the UK census we use information on population, economic activity and employ-

⁵Small opencast mines, numbering less than 100, are not included in our data set. This is mainly because we do not have any information on the location of these mines. However, we expect the importance of these mines to be negligibly small as the average number of employees is less than 10 miners, adding up to less than 1000 miners in total.

ment across industries.⁶ Data is reported down to the ward level. Unfortunately, borders of wards are redrawn frequently and we have been advised against merging these entities over time. Thus, we conduct our analysis on the district level. Also, we cannot merge districts between 1971 and 1981 due to fundamental changes in the UK's local government structure following the Local Government Act 1974. Merging the remaining periods (1981, 1991, 2001 and 2011) required some adjustments due to changes in district borders. In the case where the border of a district changes over our sample period, the districts are merged to a level which makes entities comparable over time. In 2011 this adjustment reduces the number of districts from 348 to 339. Thus, our data set consists of 339 cross-sections and 4 periods.

Population above the age of 16 is differentiated in being economically active and being economically inactive. The economically active consist of those employed and those unemployed⁷. Employed can be further disaggregated into employees and the self-employed or in full-time employment and part-time employment. Unfortunately, part-time and full-time employment among self-employed has only been reported since 2001. Thus, we treat those who reported to be part-time self-employed as in full-time employment in every year of the census. The economically inactive are differentiated in four categories: students, retired, permanently sick, and other⁸.⁹

We are able to differentiate between employment in 5 sectors (and a residual sector captur-

⁶All variables are disaggregated in two dimensions: sex and age. To construct homogenous age bins for all our indicators across time we impose the following structure because the length of the bins differed across indicators: 0-15, 16-29, 30-44, 45-59, 60+. In the case of economic activity and employment across industries, the first bin (0-15) is not available. In some cases the construction of age bins required the assumption of a uniform distribution within a bin. For example, the available age groups for the total number of employed males between 30 and 59 in 2001 is 30-39, 40-49 and 50-59. Thus, the age bin 45-59 was calculated by premultiplying 40-49 with 0.5 and adding 50-59. Similar adjustments have been required for other variables.

⁷In 1981 unemployment was constructed from two individual variables; those who reported to seeking for a job and those who reported to be temporarily sick. In 2001 the number of individuals who has been on a government scheme is reported. We treat individuals on a government scheme as unemployed.

⁸Since 2001 an additional group has been introduced indicating those who take care of family and household members. We add this to other economically inactive.

⁹Since 2001 full-time students are additionally reported to be economically active if applicable. We are forced to treat these differently in 2001 and 2011. In 2001 we add all those who reported to be full-time students and economically active to part-time employees. In 2011 the number of full-time students economically active is not reported explicitly and, instead, is added to the individual categories of economic activity. Thus, we do not observe the number of full-time students added to the individual categories of economic activity. However, it is unlikely that this will bias our results as in 2001 the number of students reported to be economically active was relatively small relative to the total number of economically active individuals, on average around 3.5%.

ing employment in "other services"): primary sector¹⁰ (agriculture, forestry, fishing, mining, energy and water supply), manufacturing sector, distribution and catering, transport and construction.¹¹ Due to the availability of data in 1981 the remaining sectors in 1991, 2001 and 2011 are aggregated to one residual variable capturing other activities in the tertiary sector.

Means of the main variables in mining and non-mining districts are presented in Table 5.6 in the Appendix for the first and the last year of our sample (see section on Identification and Estimation for a definition of our sample).

We also collect data on the EU's and the UK's governments' expenditure on regional preferential assistance to industry. The data is provided by the Office of National Statistics in an annual publication on Regional Trends. The data is disaggregated to the highest tier of sub-national division in the UK (NUTS 1 areas) - which in our sample results in 9 regions.

Data on regional assistance to industry from the UK government covers the period 1972-2003. The allocation of funds from the EU to individual regions is reported from 1975-2006. As both the UK and the EU are still actively allocating funds (the latter much more than the former) to assisted areas, we treat the remaining years as missing. Also, data on the allocation of funds from the EU is not reported on the sub-national level between 1989-1990 and 1991-1993. To resolve this problem we assume that the flows of funds are persistent and extrapolate. This does not appear to be a strong assumption. Firstly, because we observe a strong persistence of fund flows in the data for the years available. Secondly, because several authors have reported that whether a region is or is not qualified for assistance is surprisingly stable over time, what by the way calls into question the very effectiveness of regional policies (Champion and Townsend 1990).

Data for EU funds was not always available on the yearly level, but was reported cumu-

¹⁰To allow for comparability we merge the agricultural sector, fishing, forestry, mining sector and energy and water supply every year because employment in this sector is only reported as an aggregate in 2011. Note that the variable includes employment in the mining sector.

¹¹Employment across industries is reported as a random sample of 10% of the census before 2001. Thus, in 1981 and 1991 we premultiply variables reported in these categories with a factor of 10.

lately for several years (in 1980 for 1975-1980, and in 1988 for 1981-1988). This is not a serious drawback. Due to the nature of our data, the UK census being conducted every 10 years, we construct 10 years aggregates of funds flowing into assisted regions. These are designed to capture the total financial support received by assisted regions in a period of 10 years *before* the year of the census.¹²

Information on the amounts allocated by the EU are reported in European Currency Units up to 1997. Thus, we use the exchange rates of the years for which the cumulative funds are reported when converting the data into British Pounds. Before constructing the 10 years aggregates we use the UK CPI to generate real values (1994 prices). Finally, the aggregates are constructed and summed for the UK's government's expenditure and funds allocated from the EU.

5.5 Identification and Estimation

We exploit two sources of variation: the location of mines and the closure of these mines over time. The location of active mines in 1981 is presented in Figure 5.2. Coal mines were predominately located in the North East of England, in the Midlands, and South of Wales. By 2011 most of the mines had been closed. The dramatic decrease in the number of mines and number of miners is depicted in Figure 5.3. Location of the active mines for the individual census years is presented in Figure 5.5 in the Appendix.

Our identification strategy is a differences-in-differences. Our treatment group consists of districts which had an active coal mine in 1981. Our control group consists of districts without an active coal mine in 1981. Our treatment is the number of coal mines closed since 1981. In our favourite specification we restrict our sample to districts within 40 miles and exclude metropolitan boroughs and greater London (see Figure 5.2). Our results are robust

¹²For example, the total amount of annual funds between 1971-1980 received by individual regions is aggregated and controlled for in the equation of year 1981; similarly the total amount of funds between 1981 - 1990 is aggregated and controlled for in the equation of year 1991, etc.

Figure 5.2: Treatment and Control Group

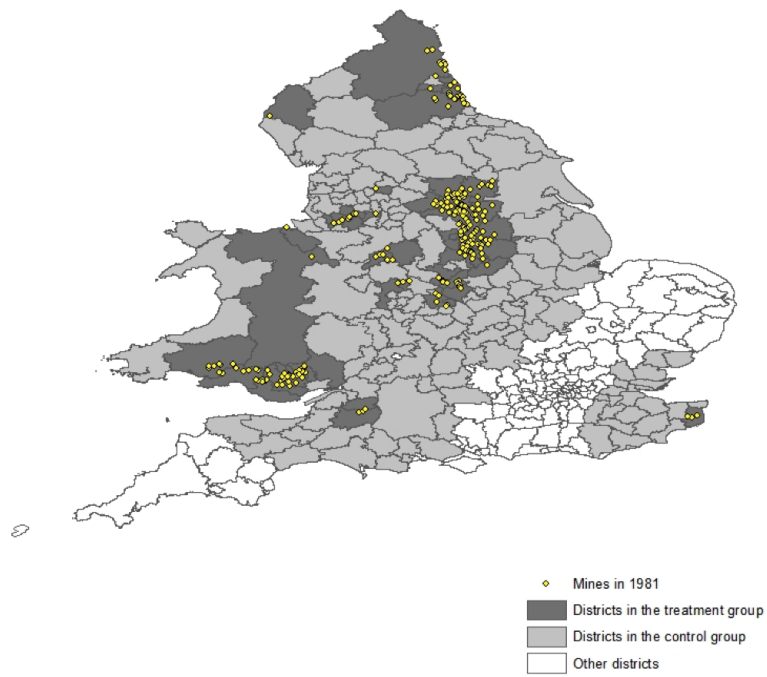
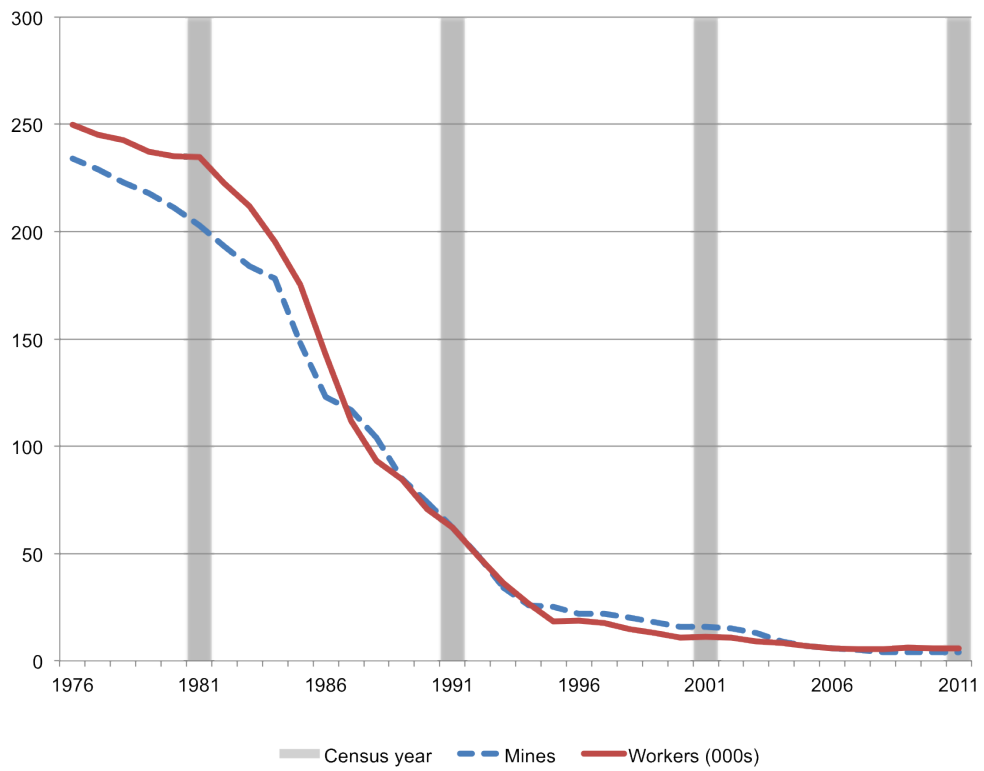


Figure 5.3: Mines and Miners



to a change of the control group. Formally, we estimate the following specification:

$$y_{it} = \beta MinesClosure_{it} + \gamma_t + u_i + \epsilon_{it} \quad (5.3)$$

y_{it} is the outcome variable of district i in period t and differs depending on our specification. *MinesClosure* is the number of mines closed since 1981. Thus, the variable is equal to zero if there were no active mines in a district in 1981; γ and u are time and district fixed effect, respectively; ϵ is the district specific error. Assuming that the errors may be correlated over time within a district and across districts within the same county in the same period, we cluster on the highest possible level (Angrist and Pischke 2009). To evaluate the persistence of these relationships we estimate the following specification:

$$y_{it} = \sum_{j=1991}^{2011} \beta_j D_i T_j + \gamma_t + u_i + \epsilon_{it} \quad (5.4)$$

D_i is equal to 1 if at least one active coal mine was located in the district in 1981. T_j is a time dummy which is equal to 1 if $t = j$. Note that this specification does not take into account the intensity of the treatment: a district in which only 1 mines was closed is treated identically to a district in which several mines were closed.

5.6 Results

The exploited variation in our experiment comes from a negative labour demand shock in the mining sector (see Figure 5.3). The specific feature of such a shock is the predominance of men being laid off. This is because mining is traditionally a male-dominated activity. The dominance of males in the mining industry is reflected in the different gender ratio in mining and non-mining districts. As can be seen in Table 5.6 in the Appendix the share of females is significantly smaller in 1981 but not so in 2011 after the closer of coal mines. Additionally, employment in the primary sector is larger in mining districts and relative to non-mining districts much more dominated by men in 1981 but not in 2011. The lay-offs in the mining industry are reflected in the results of Table 5.1. In the first three columns our explanatory

Table 5.1: Dependent variable: Employment in primary sector

	Total	Male	Female	Total	Male	Female
	(1)	(2)	(3)	(4)	(5)	(6)
Nr of mines closed within the district since 1981	-973.507 (123.091)***	-960.816 (128.627)***	-12.691 (12.116)			
Nr of mine workers laid off since 1981				-1.098 (0.08)***	-1.079 (0.076)***	-0.019 (0.021)
Obs.	664	664	664	664	664	664
R ²	0.522	0.586	0.227	0.554	0.628	0.227

Notes: Standard errors adjusted for clustering at the county level are in parentheses : *** p<0.01, ** p<0.05, * p<0.1. All regressions are estimates using OLS including district specific and time fixed effects. Sample includes districts within 40 miles of an active mine in 1981 and outside metropolitan areas.

Table 5.2: Dependent variable: Population and Employment

	ln(population)	ln(working)	participation %	unemp. %
	(1)	(2)	(3)	(4)
Nr of mines closed within the district since 1981	-0.011 (0.003)***	-0.017 (0.004)***	-.362 (0.089)***	-.100 (0.11)
Obs.	664	664	664	664
R ²	0.303	0.541	0.737	0.761

Notes: Standard errors adjusted for clustering at the county level are in parentheses : *** p<0.01, ** p<0.05, * p<0.1. All regressions are estimates using OLS including district specific and time fixed effects. Sample includes districts within 40 miles of an active mine in 1981 and outside metropolitan areas.

variable is the number of mines closed. In the last three columns our explanatory variable is the number of miners laid off. The dependent variable is the number of workers employed in the primary sector which includes mining. Three things are noteworthy. First, the coefficient in columns 3 and 6 is, as expected, very small and insignificant. Thus, female employment in the primary sector is not associated with coal mine closures. This suggests that coal mine closures indeed only affected males but not female workers in the primary sector. Second, the coefficient in column 5 suggests that laying off 1 miner is associated with a reduction of 1 male worker in the primary sector. Third, the closure of a coal mine is associated with a reduction of approximately 970 workers (column 1). This is consistent with our descriptive statistics which suggest that the average active mine in our sample employed 970 workers.

As discussed in the theoretical section, miners have several options after being laid off. They can migrate, they can exit the labour force, or they can look for jobs in other industries

Table 5.3: Dependent variable: Employment across Industries

	<u>ln(work in agri.&energy)</u>	<u>ln(work in manuf.)</u>	<u>ln(work in service)</u>
	(1)	(2)	(3)
Nr of mines closed within the district since 1981	-0.155 (0.025)***	0.008 (0.008)	0.0002 (0.007)
Obs.	664	664	664
R ²	0.704	0.663	0.833

Notes: Standard errors adjusted for clustering at the county level are in parentheses : *** p<0.01, ** p<0.05, * p<0.1. All regressions are estimates using OLS including district specific and time fixed effects. Sample includes districts within 40 miles of an active mine in 1981 and outside metropolitan areas. Services include transportation, distribution and catering, and construction.

and remain unemployed as long as they are unsuccessful. Tables 5.2 and 5.3 shed light on what happened to miners. First, column 1 in Table 5.2 suggests that the population in districts which experienced a mine closure decreased. It appears intuitive that this is the result of out-migration. A further disaggregation, not reported here, suggests that this is mainly driven by males and females between ages 45 and 59. This implies, that in particular, people with grown up children migrated. This weakens an alternative argument: the relative decrease in population might have been caused by reduced fertility or a higher death rate. The coefficient suggests that the closure of 1 mine is associated with a relative decrease in the district's population of 1%. Second, the coefficient in column 3 suggests that the participation rate decreased in the districts affected. This result has been discussed in the theoretical section and is completely consistent with a negative labour demand shock. The idea that miners coped with being laid off by exiting the labour force has already been documented (Beatty and Fothergill 1996; Beatty et al. 2007). The coefficient implies that the closure of 1 mine is associated with a 0.36 percentage point relative decrease in the participation rate. Finally, note that the coefficient in column 4 is insignificant and negative. This is not an intuitive result given the huge lay-off of miners. However, it simply suggests that miners migrated, exited the labour force or found employment in other industries rather than staying unemployed.

The results of employment across industries is reported in Table 5.3. As shown in Table 5.1, the positive coefficient in column 1 of Table 5.3 is the result of an identity. More interesting are the results in columns 2 and 3. According to the theoretical framework we would

expect in the presence of a reverse Dutch Disease an increase in manufacturing employment. On the other hand, a decrease in manufacturing employment would suggest that the loss in agglomeration due to the closure of coal mines dominated the effect on manufacturing employment. Similarly, our model suggests that employment in the service sector should decrease following a negative labour demand shock due to a drop in spending. However, this effect could be compensated by reduced wages in combination with unemployment benefits and generous lay-off compensations received by miners. In any case, the small and insignificant coefficients suggest some opposing forces and potential heterogeneous effects which are explored in the next subsection.

5.6.1 Substitution of Female Workers

Our data allows us to differentiate how coal-mine closure is associated with male and female employment. Interestingly, results in Table 5.4 suggest that, despite the fact that predominately males have been laid off, females are also affected; both the number of females working and the participation rate of females decreased in the affected regions. Note that the means in Table 5.6 (see Appendix) suggest that female participation in 1981, before the closure of coal mines, was not significantly different between mining and non-mining districts (34%) but significantly smaller in 2011 (47.3% versus 46.6%).

As suggested in the theoretical section, this could be driven by males and females being imperfect substitutes. Thus, a negative labour demand shock to males decreases the cost of male labour. Firms adjust and substitute the relatively more expensive input. This is more formally tested in Table 5.5. There are several things to notice. First, mine closure is positively associated with the number of males in manufacturing. The same is true for employment in services, even though the coefficient is small and insignificant. On the other hand, females employment in both industries is negatively associated with coal mine closures and, as before, significant in the case of manufacturing but insignificant in the case of services. At the first glance, these coefficients suggest that males *replaced* females. Note that the reduction is not trivial. The share of females in manufacturing was significantly larger in mining districts in

1981 (32.2% versus 26.7%). By 2011 the situation reversed, the share of females in manufacturing was larger in non-mining districts (23.8% versus 24.6%). A similar development, albeit not as strong, took place in services.

The results so far are consistent with our story that males replaced females in other industries (in particular manufacturing) after being laid off. However, they are also consistent with a negative female supply shock which is correlated with coal-mine closures. This could be the case if females predominantly migrated from districts which had been affected by coal-mine closures. If this is the case we should observe a decrease in the female population. On the other hand, if this change is driven by an altered industry demand, females may have remained within the same district but exited the labour force. Thus, in Table 5.7 (see Appendix) we control for the female share in the total population and the female participation rate. The results suggest that the former does little to explain the number of females in manufacturing. On the other hand, after controlling for the participation rate, the coefficient on the number of mines closed drops and becomes insignificant. We interpret this as evidence for our demand story.

We also find that these results are persistent. The results in Figure 5.4 represent the coefficients β_j from specification 5.4. They reflect the average difference between mining and non-mining districts in a given year relative to 1981. The results suggest that the population level and the participation rate of the districts affected are on average significantly below the levels of our control group up to 2011, even though, the participation rate shows some sign of recovery. The same applies to the number of females and the proportion of females in manufacturing. The results point to a persistent substitution of females in the manufacturing sector in the mining districts. These results are consistent with recent work documenting negative and persistent effects of resource booms on mining towns in the US (Jacobsen and Parker 2013).

Additionally, evidence for our demand story can be derived based on the results from Beatty and Fothergill (1996) and Beatty et al. (2007). They suggest that most miners coped

Table 5.4: Dependent variable: Employment of Males and Females

	ln(fem. work) (1)	ln(male work) (2)	fem. partic. % (3)	male partic. % (4)
Nr of mines closed within the district since 1981	-0.015 (0.004)***	-0.019 (0.004)***	-0.303 (0.104)***	-0.383 (0.084)***
Obs.	664	664	664	664
R ²	0.759	0.24	0.934	0.811

Notes: Standard errors adjusted for clustering at the county level are in parentheses : *** p<0.01, ** p<0.05, * p<0.1. All regressions are estimates using OLS including district specific and time fixed effects. Sample includes districts within 40 miles of an active mine in 1981 and outside metropolitan areas.

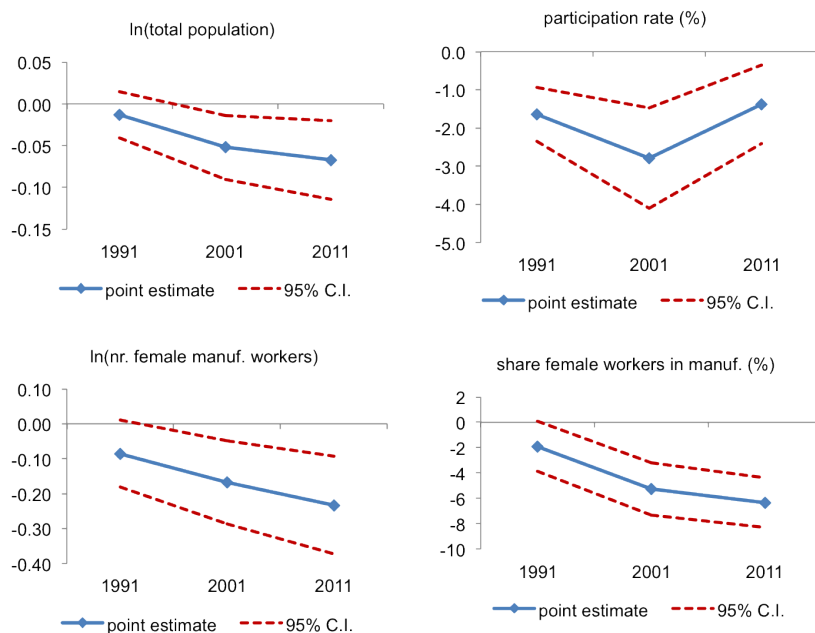
Table 5.5: Dependent variable: Employment of Males and Females across Industries

	ln(fem. manuf.) (1)	ln(male manuf.) (2)	ln(fem. serv.) (3)	ln(male serv.) (4)
Nr of mines closed within the district since 1981	-0.027 (0.009)***	0.024 (0.01)**	-0.007 (0.008)	0.003 (0.009)
Obs.	664	664	664	664
R ²	0.701	0.615	0.711	0.732

Notes: Standard errors adjusted for clustering at the county level are in parentheses : *** p<0.01, ** p<0.05, * p<0.1. All regressions are estimates using OLS including district specific and time fixed effects. Sample includes districts within 40 miles of an active mine in 1981 and outside metropolitan areas. Services include transportation, distribution and catering, and construction.

with being laid off by exiting the labour force: they either retired or registered as permanently sick. If our results are driven by a substitution of females in the labour force we would expect that females coped with being substituted in a similar fashion. To test for that we estimate specification 5.4 with the outcome variable being either the logged number of females retired or permanently sick, or the proportion of females retired or permanently sick among females over 15. The results are presented in Table 5.8 in the Appendix. First, column 2 and 4 confirm the results presented by Beatty et al. (2007): the number of men retired or permanently sick in districts which had at least one active coal mine in 1981 increased relative to districts without an active coal mine since 1981. The results in column 1 and 3 suggest that this relationship is also present for females. Interestingly, the results also suggest that the relationship is stronger and even more persistent.

Figure 5.4: Persistence



5.6.2 Robustness

Spatial spillover: One might speculate that the closure of coal mines affects neighboring districts. Surprisingly, this does not appear to be the case. Results in Figure 5.6 (see Appendix) suggest no significant relation between any of our main variables of interest and coal mine closures. This suggests a very localized effect of coal-mine closures on all of our main outcome variables. Nevertheless, we reproduced our main results using Conley standard errors (Conley 1999) and the significance of our results is not affected.

Government Expenditure: The UK has been subject to a dynamic history of policies targeting local economic development since 1945 (see, for example, Hall (2007)). Obviously, we cannot abstract from regional economic development policies when thinking about impacts on local labour markets, mainly because these are endogenous to coal-mine closures. Usually, these policies targeted a reduction in structural unemployment in less developed areas, and there is some evidence that they were successful in doing so up to the 1970s (Moore and Rhodes 1973; Moore et al. 1986). Since the mid 1970s there have been serious attempts to rely more on free market forces and reduce regional economic support (Hall 2007; Beatty

et al. 2007; Broadberry and Leunig 2013). These attempts have been counterbalanced by the European Union's regional policy goal to achieve economic and social cohesion since the 1980s. Additionally, coalfields have been targeted by several specific regeneration programs since 1985 (Beatty et al. 2007). Despite this, several studies agree that regional policies had barely any of the intended effects on the geography of *manufacturing employment* in the UK since the 1970s (Wren 1996; Beynon and Hudson 1999; Wren and Taylor 1999; Hall 2007; Broadberry and Leunig 2013). Nevertheless, this is an important confounding factor which we have to control for.

Optimally, we would have data on funds allocated on the district level for every individual project involved in the regeneration of regions affected by coal-mine closure. Unfortunately, we were not able to get hold of such data. However, we were able to collect data on financial flows disaggregated down to the highest tier of sub-national division in the UK (NUTS 1 areas) and two most important sources of funding (Champion and Townsend 1990; Beatty et al. 2007). The EU regional development grants, which were introduced in 1975. And UK government expenditure on the economic development of assisted areas, which has its roots in the 1930s. As presented in Tables 5.9 and 5.10 (see Appendix), our results are not significantly affected after the inclusion of the constructed variable (see section Data for a description of how the variable is constructed). Alternatively, we use region year fixed effects to capture non-linear trends on the regional level. Even though some coefficients turn insignificant they still have the same sign and a similar magnitude.

5.7 Conclusion

We use the closure of coal mines in the UK since 1981 as a treatment in our quasi natural experiment to study labour markets in districts affected by the coal bust. To do that we construct a new data set containing information on the location of coal mines, the number of miners active within individual coal mines and the year of closure from 1981 onwards. We spatially link this data set to the UK census for the period 1981-2011. Our treatment group consists

of districts which contained at least one active mine in 1981. Our control group consists of districts without a coal mine, but within 40 miles of an active mine (excluding metropolitan areas).

We find that districts affected by the coal bust experienced a rapid and permanent relative decrease in population. But we do not observe an increase in unemployment in districts affected by coal mine closures. This is because individuals were willing to migrate to equalize employment opportunities, exit the labour force or find a job in other industries. On the aggregated level (males and females) we find evidence for the first two channels but do not find any significant relationship with coal mine closures and employment in manufacturing or services. Disaggregating the data by gender suggests that the closure of coal mines is positively associated with an increased employment of males and a decreased employment of females in manufacturing. Moreover, our results suggest that a larger number of both, males and *females*, exited the labour force in the effected regions, despite the closure of mines being a negative labour demand shock to males only. We interpret that as evidence that men being laid off in the mining sector substituted females in other sectors who subsequently exited the labour force. Further tests suggest that significantly more females registered as permanently sick or retired in districts affected by coal mine closures. Such a transmission of a negative labour demand shock is quite intuitive if we consider males and females to be imperfect substitutes.

5.8 Appendix

Figure 5.5: Location of Active Mines

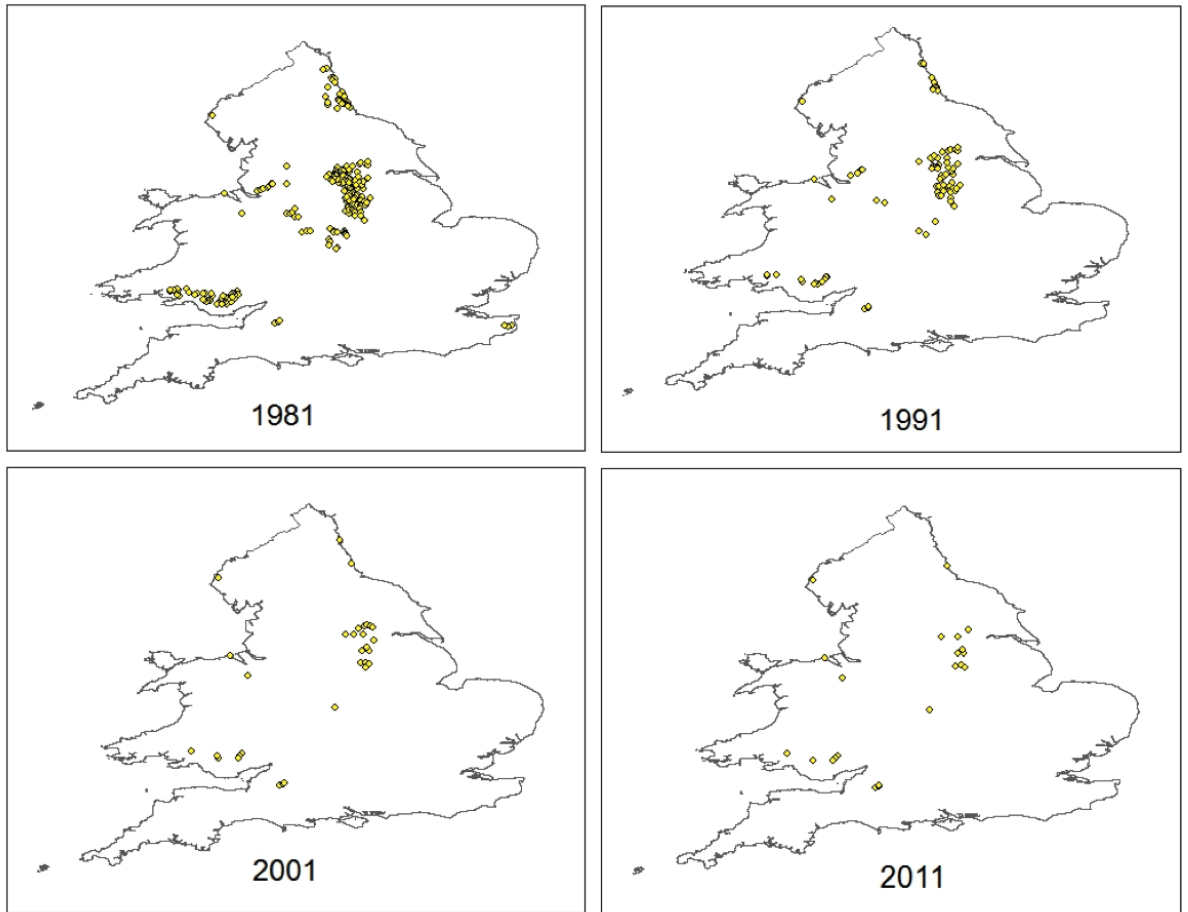


Table 5.6: Descriptives

Variables	Mean Mining (1981)	Mean Mining (2011)	Mean Non-Mining (1981)	Mean Non-Mining (2011)
Population (000)	153.2	168.8	112.7	133.6
Females (000)	78.2	85.7	57.6	67.9
Females/Population*100	50.8	50.7	51.1	50.8
Labour Force (000)	70.6	84.1	52.4	68.9
Labour Force/Pop.*100	46.5	50.5	46.4	51.4
Fem. in Labour Force/Fem.*100	34.5	46.6	34.4	47.3
Working (000)	62.6	77.9	47.3	64.4
Unemployed/Labour Force*100	9.7	7	8.6	6.3
Primary (000)	6.7	2.4	2.8	1.9
Primary /Working*100	11.6	3.1	6.5	3.3
Fem. in Primary/Primary*100	10.1	23	17.3	23.1
Manufacturing (000)	18.7	9.2	13.6	6.8
Manufacturing/Working*100	31.4	12.7	28.5	10.8
Fem. in Manuf./Manuf.*100	32.2	23.8	26.7	24.6
Services (000)	18.9	29	15.5	24.4
Services/Working*100	29.5	37.8	32.8	37.9
Fem. in Services/Services*100	38.1	37	36.4	36.8

Figure 5.6: Spatial spillover

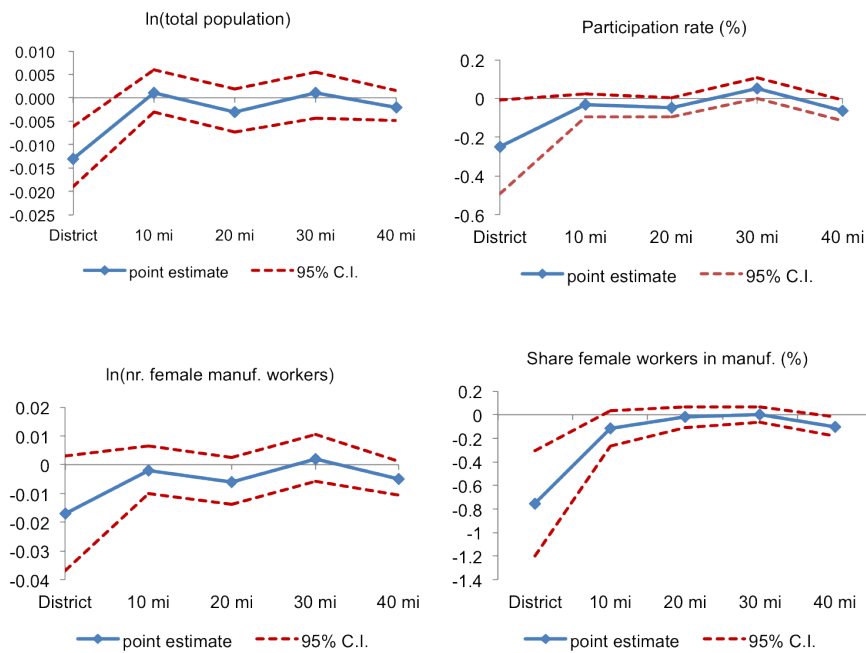


Table 5.7: Dependent variable: Females in Manufacturing

	ln(fem. manuf.)	ln(fem. manuf.)	ln(fem. manuf.)	ln(fem. manuf.)
	(1)	(2)	(3)	(4)
Nr of mines closed within the district since 1981	-0.027 (0.009)***	-0.024 (0.01)**	-0.010 (0.009)	-0.010 (0.01)
Females (%)		-6.786 (7.512)		-2.244 (5.626)
Females participation (%)			0.055 (0.007)***	0.055 (0.006)***
Obs.	664	664	664	664
R^2	0.701	0.705	0.772	0.772

Notes: Standard errors adjusted for clustering at the county level are in parentheses : *** p<0.01, ** p<0.05, * p<0.1. All regressions are estimates using OLS including district specific and time fixed effects. Sample includes districts within 40 miles of an active mine in 1981 and outside metropolitan areas.

Table 5.8: Dependent variable: Permanently Sick and Retired

	share of fem.	share of males	ln(fem.)	ln(male)
	(1)	(2)	(3)	(4)
β_{1991}	0.014 (0.003)***	0.017 (0.003)***	0.095 (0.037)**	0.04 (0.014)***
β_{2001}	0.028 (0.005)***	0.017 (0.004)***	0.128 (0.053)**	0.026 (0.022)
β_{2011}	0.018 (0.007)***	0.007 (0.005)	0.056 (0.062)	-0.064 (0.029)**
Obs.	664	664	664	664
R^2	0.947	0.816	0.944	0.843

Notes: Standard errors adjusted for clustering at the county level are in parentheses : *** p<0.01, ** p<0.05, * p<0.1. All regressions are estimates using OLS including district specific and time fixed effects. Sample includes districts within 40 miles of an active mine in 1981 and outside metropolitan areas.

Table 5.9: Dependent variable: Population and Participation

	ln(pop.) (1)	partic. % (2)	ln(pop.) (3)	partic. % (4)	ln(pop.) (5)	partic. % (6)
Nr of mines closed within the district since 1981	-.011 (0.003)***	-.362 (0.089)***	-.009 (0.003)***	-.321 (0.087)***	-.009 (0.004)**	-.116 (0.099)
EU and UK expenditure on assisting economic development			0.011 (0.004)***	0.228 (0.114)**		
Region-Year FE	N	N	N	N	Y	Y
Obs.	664	664	664	664	664	664
R ²	0.303	0.737	0.315	0.742	0.345	0.814

Notes: Standard errors adjusted for clustering at the county level are in parentheses : *** p<0.01, ** p<0.05, * p<0.1. All regressions are estimates using OLS including district specific and time fixed effects. Sample includes districts within 40 miles of an active mine in 1981 and outside metropolitan areas.

Table 5.10: Dependent variable: Employment in Manufacturing

	ln(fem.) (1)	ln(male) (2)	ln(fem.) (3)	ln(male) (4)	ln(fem.) (5)	ln(male) (6)
Nr of mines closed within the district since 1981	-.027 (0.009)***	0.024 (0.01)**	-.025 (0.009)***	0.022 (0.01)**	-.035 (0.014)**	0.019 (0.012)
EU and UK expenditure on assisting economic development			0.013 (0.01)	-.011 (0.009)		
Region-Year FE	N	N	N	N	Y	Y
Obs.	664	664	664	664	664	664
R ²	0.701	0.615	0.702	0.616	0.723	0.664

Notes: Standard errors adjusted for clustering at the county level are in parentheses : *** p<0.01, ** p<0.05, * p<0.1. All regressions are estimates using OLS including district specific and time fixed effects. Sample includes districts within 40 miles of an active mine in 1981 and outside metropolitan areas.

Chapter 6

Modelling Shocks in the Upstream Sector of the Oil and Gas Industry

6.1 Introduction

In 1931 Hotelling suggested a way of thinking about the optimal extraction path and the corresponding price path of an exhaustible resource. Hotelling (1931) suggested thinking of natural resource extraction as a cake-eating problem - whatever is extracted today cannot be extracted tomorrow. As a result the extraction of the resource is driven by an arbitrage condition - extract and reinvest if returns from investment are higher than keeping the resource in the ground and vice versa. Based on this reasoning, the oil price should be determined by the rate of interest.

More than 80 years later the economic profession is still struggling to understand the crude oil price and the shocks driving it, and is still questioning the applicability of Hotelling's rule (Adelman 1962; Adelman 1990; Pindyck 1999; Hamilton 2008; Kilian 2009; Venables 2011; Anderson et al. 2014). Our data allows us to contribute to our understanding of the determinants of drilling costs and their impact on the real price of oil. We use micro data from Wood MacKenzie on the number of exploration wells drilled and costs of drilling. Focusing on the information provided for the biggest 25 oil and gas companies we construct two quarterly time series capturing (i) the total number of exploration wells drilled by these companies

and (ii) the average cost of drilling these wells. We use the constructed time series in combination with the real price of crude oil to estimate a three dimensional structural VAR model.

The proposed structural model of the **upstream sector** allows us to decompose the variation in the reduced form errors of the estimated VAR into three structural shocks. To identify the shocks we assume a recursive structure as in Kilian (2009). The first structural shock is an *oil price demand shock* which is defined as an unpredictable innovations to the oil price. To identify the shock we assume that the oil price is predetermined to drilling activity and costs of drilling. This assumption is plausible because it takes more than three months, and typically more than a year, following the drilling of an exploration well before the global supply of oil and gas can be affected. Moreover, expectations of the future oil price are unlikely to be formed contemporaneously because it takes usually more than a year before reasonable estimates of reserves and drilling costs are available.

The remaining variation in the errors - after accounting for the variation in the oil price - is decomposed into a *supply shock* and a *drilling specific demand shock*. We think of a supply shock being driven by technological changes, e.g. the development of horizontal drilling. The demand shock is referred to as drilling specific to emphasize that the shock originates in the upstream sector and is *not* driven by fluctuations in the oil price. We think of a drilling specific demand shock as being mainly driven by regulatory changes: e.g. increase in offshore drilling following the expansion of area available for leases, the nationalization of an oil and gas industry restricting access to specific regions, an increase in taxes or tougher safety regulations. We disentangle the two shocks by assuming that the number of exploration wells which can be drilled in the short run is fixed (vertical supply curve). Thus, only supply shocks can affect the number of wells drilled and costs of drilling contemporaneously. We justify this assumption by arguing that the exploration of new territories is often associated with the necessity to utilize and transport rigs, and contract, maybe even educate, a new crew to operate the rig in a new environment (Yergin 2011).

We have three main results. First, a 1% increase in drilling activity increases costs of

drilling by 0.5-1% following an oil price driven demand shock. This implies an average long run (12 quarters) price elasticity of aggregate supply of 1-2. We also observe the existence of significant heterogeneity in the slope of the supply curve. In particular, our results suggest that the supply curve is steeper in countries with a lower GDP per capita. Second, supply shocks do not have a significant effect on costs, suggesting a flat demand curve in the upstream sector. Third, shocks to drilling costs do *not* have a permanent effect on the price of oil.

Thus, we confirm the theoretically expected fact that the extraction of natural resources is an increasing costs industry. More interestingly, shocks in the upstream sector do not appear to be important in determining the price of oil in the long run, calling into question the theoretically expected link between extraction costs and the price of oil (Krautkraemer 1998). Partly, this result may be explained by the endogenous expansion of crude oil production as a reaction to an increase in drilling costs (Kilian 2009).

Our work builds on a vast literature starting with Hotelling (1931). Many theoretical adjustments have been made to the original model since then. More importantly, most assumptions and predictions of the model have been tested and rejected. A literature review on theoretical advancements and the empirical tests is provided by Krautkraemer (1998). A slightly earlier but more comprehensive summary of the economics of petroleum supply comes from Adelman (1990). The most comprehensive summary of the history of oil and gas extraction is provided by Yergin (2011).

Our paper is closely related to Anderson et al. (2014). Using data from Texas on drilling activity and rig rents, they present evidence that drilling activity and drilling costs significantly respond to a change in the oil price. On the other hand, they are not able to find any significant relationship of oil price changes and the contemporaneous extraction of oil. They use these results to motivate a theoretical model in which drilling activity is at the centre of deriving an optimal extraction path. In doing so, they are able to rationalize their empirical findings. Our work differs from their contribution in two main aspects. First, and most impor-

tantly, their main contribution is theoretical whereas our focus is on the causal identification and estimation of shocks. Second, we use data on drilling and costs of drilling covering the whole world, whereas their data is constraint to activity and rig rents in Texas.

In our approach to model the upstream sector we are very closely following Kilian (2009) who uses a three-dimensional structural VAR and short-run restrictions to understand the different nature of shocks driving the oil price. As opposed to our approach, he models the world market for crude oil rather than the upstream sector of the oil and gas industry. To do this he uses world oil production to capture supply shocks. To model demand shocks for industrial commodities, he uses dry cargo single voyage ocean freight rates. After accounting for shifts in demand and supply, he argues that the residual variation in the oil price is driven by precautionary demand shocks rather than by supply shocks as has been previously believed.

This paper is structured as follows. In the next section we discuss the institutional framework of the upstream sector in the oil and gas industry. A simple theoretical framework is presented in the third section. In the fourth section we describe the data. In the fifth section we discuss our identification and estimation strategy. In the subsequent section our results are presented before we conclude.

6.2 Upstream Sector in the Oil and Gas Industry

Oil and gas reserves are located beneath the Earth's surface. The reserves' location is typically referred to as an oil and gas field. The mission of the oil and gas company is to locate fields and then extract, transport, process and eventually sell the product. These steps are typically subsumed in three sectors: an upstream sector, which involves locating and extracting the product; a midstream sector, which mainly involves the transportation and storage of the product; a downstream sector, which involves the processing, distribution and selling of the final product. In what follows we focus on the **upstream sector** as we are interested in the determinants of drilling costs and their interaction with the oil price.

In order to operate in a field the company requires a licence from the owner of the mineral rights. Depending on the circumstances this can be an individual, an institution or a state. Usually, several companies are simultaneously operating in the same field. Sometimes they are competing for the extraction of the product which results in common pool problems - most famously seen during the first oil boom in Pennsylvania (Yergin 2011). And sometimes they collaborate by creating consortia which often involve the state as an important agent in the project. Partly as a consequence of the variety of work relationships the type of licences allocated to companies often varies and is frequently determined on a bilateral basis involving prolonged negotiations to agree on a profit sharing scheme.

Reserves are typically buried under many layers of rock and in the case of offshore drilling these layers are covered by water. Thus, drilling rigs represent the core of the upstream sector as they are essential to penetrate the layers of rock to access the product.¹ Once the well is drilled the drilling rig is no longer required and can be moved. This very fact explains why companies in the oil and gas sector are not vertically integrated and why drilling is typically outsourced (Kellogg 2011). Drilling activity of any producer fluctuates with oil production outcomes from recently drilled wells and the firm's success in finding new fields in a particular location. A successful exploration well drilled attracts subsequent drilling from the own or a competing company, and a dry hole does not. Outsourcing drilling rigs reduces the overall capacity requirements of rigs and obviously greatly reduces transportation and mobilization costs. Despite the fact that producers outsource drilling they are responsible for identifying potential reserves, designing wells and writing drilling procedures (Kellogg 2011). This is mainly because they are better informed about the specific geological conditions. In Figure 6.6 in the Appendix the success rate of drilling a well as a share of the total number of exploration and appraisal wells drilled by companies in our sample is presented. The obvious downward trend in the share of unsuccessful wells drilled suggests technological progress in identifying productive wells.²

¹The main features of a rig are a tall derrick which stabilizes the pipe and an engine which spins the drill.

²Note that because we limit our sample to major oil and gas companies the 85% (80% if appraisal wells are dropped) success rate in 2012 might represent an upper bound if the major oil and gas companies are more

Obviously, there is a huge variation in the geology which the drill has to overcome. This is even true for holes drilled within a field (Kellogg 2011). Hence, depth of drilling and time required to drill a new hole varies considerably. As a result of the latter, at any point in time and a given level of knowledge the upstream sector is subject to increasing costs (Adelman 1962): a small output can be drawn from lower-cost and high- quality sources, but a large output has to be drawn from higher-cost low-quality sources. A change in the oil price might trigger the exploration of new, previously unprofitable, regions. Alternatively, a change in technology affecting the state of the art of drilling might reduce drilling costs and geographically expand drilling activities.

Types of wells drilled differ in their purpose. We broadly differentiate between two types of wells: exploratory wells³ and production wells. The former have the purpose of identifying new reserves. The latter are drilled to extract oil and gas. We model the upstream sector by focusing on the drilling of exploratory wells rather than on all types of wells drilled, e.g. production wells. This is done for two reasons: a theoretical and an empirical. Theoretically, exploration is most closely linked to the movement in and out of territories which have high-cost low-quality sources. This movement affects drilling costs and eventually the price of oil. On the other hand, an increased drilling activity per se will not necessarily result in higher costs of drilling if conducted in territories of well-established fields. Empirically, the oil price may be considered predetermined to activity in drilling of exploration wells - but less so in case of production wells - which is an important assumption of our identification strategy and will be discussed further below.

Typically, the contracting is initiated by the oil company requesting the drilling services and specifying the requirements of the hole drilled, e.g. the depth of the well to be drilled, specifics of geological layers, etc. Based on this information, the drilling company makes an offer by specifying the rig to be used, the crew hired and the price to be paid.⁴

successful in identifying productive wells.

³Throughout the paper we refer to exploratory and appraisal wells as exploratory wells as the latter can be considered to be part of the exploration process (Adelman 1990).

⁴See (Kellogg 2011) for a more detailed discussion of contracting between oil companies and drilling compa-

6.3 Theoretical Framework

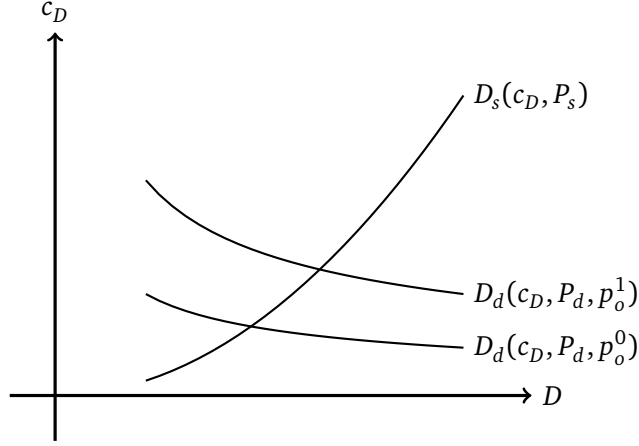
Based on the previous discussion we model the **upstream sector** of the oil and gas industry as follows. On the **demand side** the representative oil company's mission is the discovery of new reserves, R_d . The reader may think of BP, Shell or Chevron searching for new oil and gas fields. To explore the geological layers beneath the earth's surface the company has to drill new holes and, thus, demands drilling services, D_d . Using drillings as the only variable factor of production the oil company discovers new reserves according to the following production function: $R_d = A_d D_d^\alpha$. For simplicity it is assumed that there is no uncertainty in exploration drilling. Thus, every additional drilling increases the amount of reserves discovered. However, the discovery of reserves is subject to diminishing returns, $\alpha < 1$. Intuitively, this is because in every period the most promising regions are explored first. A_d is the productivity shifter but also captures other factors which are discussed below. The cost of drilling a well is denoted c_D . The price of oil, p_o , is assumed to be exogenous to the individual company. The company maximizes profits, $\pi_d = p_o R_d - D_d c_D$, by choosing the optimal number of drillings, D_d :

$$D_d = \left(\frac{c_D}{\alpha p_o A_d} \right)^{1/(\alpha-1)} \quad (6.1)$$

On the **supply side** of the **upstream sector** the representative firm provides drilling services. The reader may think of Schlumberger or Baker Hughes producing rigs and supplying drilling services to oil companies which are searching for new reserves. Engineers, E_s , are assumed to be the only variable input of production. As before the production function is assumed to be concave $D_s = A_s E_s^\beta$, with $\beta < 1$. A_s is a productivity shifter capturing technical advances. Engineers are paid the wage w . Treating c_D and w as given, firms maximize profits, $\pi_s = c_D D_d - E_s w$, by choosing the number of engineers they want to employ. With the optimal

nies.

Figure 6.1: Upstream Sector in the Oil and Gas Industry



number of engineers the supply of drilling services, D_s , is determined by

$$D_s = A_s^{1/1-\beta} \left(\frac{w}{c_D \beta} \right)^{\beta/(\beta-1)} \quad (6.2)$$

Assuming α and β to be fixed, taking logs and first differencing we can write equation 6.1 and 6.2 as

$$\begin{aligned} \frac{\Delta D_d}{D} &= -\xi \frac{\Delta c_D}{c_D} + \xi \frac{\Delta p_o}{p_o} + \xi \frac{\Delta A_d}{A_d} & \xi &= 1/(1-\alpha) \\ \frac{\Delta D_s}{D} &= \theta \frac{\Delta c_D}{c_D} + \theta \beta^{-1} \frac{\Delta A_s}{A_s} & \theta &= \beta/(1-\beta) \end{aligned} \quad (6.3)$$

From the equations in 6.3 and Figure 6.1 we see that by construction the complete variation in the number of wells drilled D and costs of drilling c_D (note that we have data on both variables) can be decomposed into three exogenous shocks: Δp_o , ΔA_s and ΔA_d . We use these shocks to interpret the structural shocks which are identified in the subsequent empirical analysis. In the next paragraph we define and interpret the shocks. The assumptions required for the identification of the shocks are discussed in the empirical section.

The first shock is an unanticipated change in the real price of oil, Δp_o . We think of this

shock as a demand shifter in the upstream sector. An increase in the oil price, *ceteris paribus*, increases the profitability of oil wells and triggers an increase in drilling activity. We refer to this shock as an *oil price demand shock*. The second shock is a change in the productivity of firms on the supply side of the upstream sector, ΔA_s . These are the firms which construct rigs and provide drilling services. Thus, we think of this shock as being driven by technological progress: the development of horizontal drilling or advances in deep sea offshore drilling. We refer to this shock as a *supply shock*. The third shock is a change in the productivity of firms on the demand side of the upstream sector, ΔA_d . These are the firms which, in their search for new reserves, demand drilling services. ΔA_d captures productivity shocks but also other factors affecting demand for drillings. Most notably, regulatory changes: an increase in offshore drilling following the expansion of area available for leases (e.g. US since the 1980s); a reduction in offshore drilling due to tougher safety regulations following an oil spill; the nationalization of an oil and gas industry restricting access to specific regions; the opening up of new regions (e.g. the collapse of the Soviet Union). We refer to this shock as a *drilling specific demand shock*. The term 'drilling specific' is chosen to emphasize that the shift in demand originates in the upstream sector and is *not* driven by fluctuations in the oil price.

Treating the shocks orthogonally and allowing the market for drilling services to clear we can derive the equilibrium response of c_D and D to each shock individually by equating D in the first and second equation of 6.3.

The response of cost of drilling a well and the number of wells drilled to an oil price shock is positive (see equation 6.4). Obviously, the response of drilling costs to a change in the oil price is smaller than 1 because D adjusts and, hence, partly absorbs the shock. If in the short run (within a quarter) D is not able to adjust, the increase in the oil price is completely captured by the change in unit cost of drilling, $\Delta c_D/c_D = \Delta p_o/p_o$. The response

to the drillings specific demand shock is equivalent by construction and is not derived here.

$$\begin{aligned}\frac{\Delta c_D}{c_D} &= \frac{\xi}{\theta + \xi} \frac{\Delta p_o}{p_o} \\ \frac{\Delta D}{D} &= \frac{\theta \xi}{\theta + \xi} \frac{\Delta p_o}{p_o}\end{aligned}\tag{6.4}$$

A supply shock affects costs negatively and activity positively (see equation 6.5). A decrease in costs of drilling and an increase in activity in the upstream sector suggests that the oil price should be affected negatively (assuming the size of discovered reservoirs and the success rate of discoveries do not change).

$$\begin{aligned}\frac{\Delta c_D}{c_D} &= -\frac{\theta \beta^{-1}}{\theta + \xi} \frac{\Delta A_s}{A_s} \\ \frac{\Delta D}{D} &= \frac{\xi \theta \beta^{-1}}{\theta + \xi} \frac{\Delta A_s}{A_s}\end{aligned}\tag{6.5}$$

6.4 Data

In the first part of this section we discuss the source and the quality of our data. In the second part we discuss the transformation of the data, before continuing with a discussion on identification and estimation in the next section.

The raw dataset was obtained from Wood Mackenzie, and compiled using a range of methods: (i) meetings with energy companies, annual reports and industry specific publications, (ii) state publications and information from public institutions, (iii) historical data, investor presentations and different types of media sources.

We use the data to construct two variables. *Drilling activity*, which is the total number of exploration and appraisal wells (henceforth exploration wells) drilled by a particular company in a particular location (onshore/offshore) in a particular quarter⁵. *Costs of drilling a*

⁵We have information on the starting day of drilling and on the completion day of drilling. We use the former because we are interested in how quickly companies in the oil and gas sector respond to an increase in the crude oil price.

metre, which represents the variable costs⁶ in \$US of drilling a metre. Unit costs of drilling a well are first transformed into real values using US CPI⁷ and then normalized by the depth of the well.

To our knowledge, this is the best source of data on drilling activities and costs in the oil and gas industry covering the whole world. Nevertheless, we are concerned about the quality of the reported data and the fact that approximately 20% of the reported cost of drilling a metre is missing. We conduct several steps to address these concerns.

First, we do not use the complete sample available to us from Wood MacKenzie. We limit the data in cross section and time dimension. In Figure 6.2 all companies used in the empirical analysis are presented. Some basic information on the selected companies can be found in Table 6.2 in the Appendix. Choosing the major companies is motivated by the idea that the information provided for and by these companies is more reliable. This is because publicly traded companies are required to be transparent. Moreover, many active interest groups watch the activity of these companies carefully (e.g. shareholders, governments, environmental activists, competitors). We have picked the 25 biggest oil and gas companies based on their production levels as reported by Forbes (Helman 2013).⁸ Restricting the analysis in this way allows us to capture the important changes in activity and costs in this sector for two reasons. In total the 25 selected companies produced more than 50% of total oil and gas production in every year of our sample. Moreover, these companies are arguably at the technological frontier and the first either to engage in a new type of drilling or the first to explore a new region of the world.

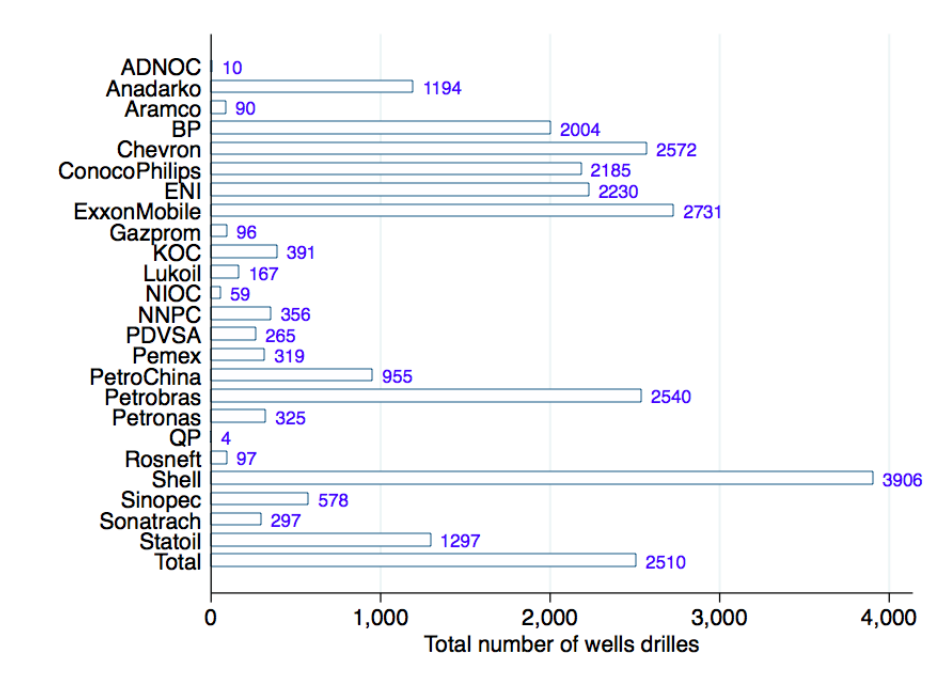
Second, we limit our data in the time dimension by starting our analysis in 1992 because the quality of data appears to decrease significantly going back in time (many more observations on depth and costs of drilling a well are missing before 1992).

⁶We note that the calculation of costs per well can vary depending on the region and the company involved. However, there is a great deal of effort from company's employees to construct comparable variable costs of drilling and exclude "back office" costs.

⁷Quarterly CPI for the US is taken from OECD statistics.

⁸We excluded Iraq due to its turbulent recent past. Instead we add Anadarko as one of the biggest private companies in the world.

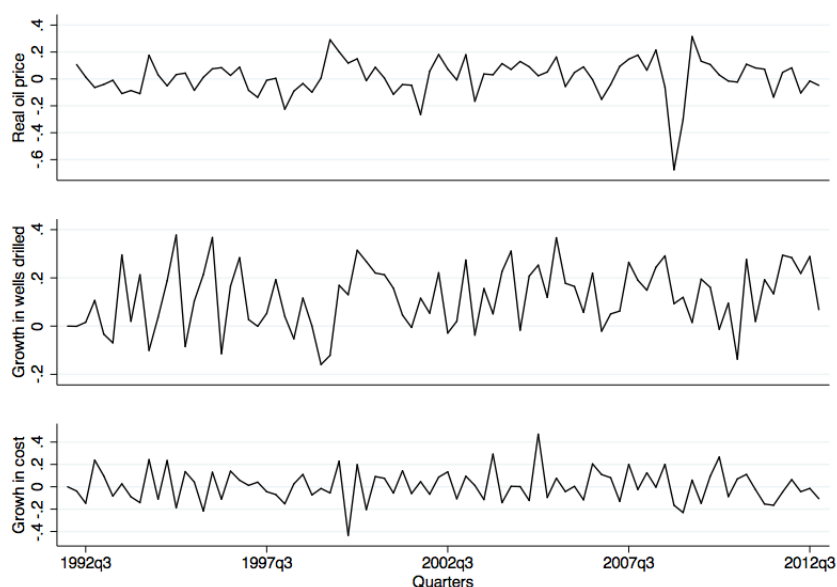
Figure 6.2: Number of wells drilled by company



Third, approximately 23% of the information on costs per metre drilled is missing.⁹ We are particularly worried about missing values which could affect a consistent estimate of the mean. This would not be a problem if information on wells drilled is missing at random. Thus, in what follows we will present some evidence that this seems to be the case. For the calculation of costs per metre we use the information on unit costs of drilling a well and the depth of the well drilled. Luckily, there are only 125 cases where information is not available for both variables. With 27178 observations in total this represents less than 0.5% of our sample. Such a small number is rather unlikely to affect our estimates of the mean significantly. In 3209 cases we are only missing information on the depth of the well drilled. And in 3283 cases we are only missing information on the costs of the well drilled. Of course, there will be a strong positive correlation between the depth of the well and the variable costs involved in drilling the well. In our sample the correlation coefficient is 0.46. Being aware of the positive correlation we can get an idea about the distribution of the missing variables by using the available information. In Figure 6.7 (see Appendix) we plot the distribution of variable costs

⁹Also, it is important to mention that we are completely missing information on US onshore drilling. Unfortunately, there is little we can do about it.

Figure 6.3: Growth of the constructed time series



in \$US (logged) for observations where the information on depth is missing, and where it is not missing. In Figure 6.8 (see Appendix) we plot the distribution of the depth of the wells for observations where costs are missing and where costs are not missing. Comparing medians and means of the corresponding distributions we cannot reject the hypothesis of no significant differences.¹⁰ The similarity of distributions and the insignificant difference of means and medians suggests that, for our purposes, we might treat the missing observations as missing at random. Thus we replace the missing values with the group specific mean when calculating the company-location-quarter specific means.¹¹

For the subsequent empirical analysis we transform the data as follows. We want to take into account location and company fixed effects in both dimensions: activity and costs.¹²

¹⁰We regress costs and depth of a well on a dummy indicating missing information on depth and costs respectively.

¹¹List-wise deletion of variables appears inappropriate because this would bias the estimate on activity downwards.

¹²We initially considered accounting for country fixed effects. For two reasons we have decided against this approach. First, the increasing sparsity of data points when disaggregating further. Second, we think of the companies as being globally integrated. Thus, variable cost (not taking into account the required negotiations over licences and the instability of countries) is driven by the technological abilities of firms, the location drilled, and the depth of the hole drilled rather than national borders.

Table 6.1: Descriptive statistics of the main variables

Variable	Mean	Median	SD	Min	Max
Δ in the real oil price	0.01	0.03	0.14	-0.68	0.31
Δ in number of wells drilled	0.12	0.12	0.13	-0.2	0.4
Δ in costs of drilling a metre	0.01	0	0.14	-0.44	0.47

Moreover, we need to take into account that these variables are not stationary. In order to do this we partly follow Kilian (2009) in his approach. We construct quarterly aggregates to capture drilling activity and costs of drilling on location¹³ and company level (sum for drilling activity and mean for costs per metre drilled). The constructed time series is then logged and first differenced to account for the individual fixed effects. Finally, the individual series are weighted according to the number of wells drilled relative to the total number of wells drilled in the particular quarter. The resulting time series are presented in the second and third panels of Figure 6.3. In the first panel of Figure 6.3 the percentage change of the real oil price is presented. The nominal monthly oil price for WTI (Cushing) is taken from EIA. The oil price is transformed into quarterly values using the arithmetic mean, and into real values using US quarterly CPI.¹⁴ Descriptive statistics of our variables are presented in Table 6.1.

6.5 Identification and Estimation

Based on the theoretical section, we set up a structural model using quarterly data consisting of $Y_t = (\Delta \ln(p_o), \Delta \ln(D), \Delta \ln(c_D))'$: the growth in the real price of oil, $\Delta \ln(p_o)$; the growth in the number of wells drilled, $\Delta \ln(D)$; the growth in costs of drilling a well (normalized by the depth of the well), $\Delta \ln(c_D)$.¹⁵ The structural VAR representation is

$$\mathbf{A}_0 \mathbf{Y}_t = \alpha + \sum_{i=1}^6 \mathbf{A}_i \mathbf{Y}_{t-i} + \varepsilon_t \quad (6.6)$$

α is a vector of constants capturing the average growth rate. \mathbf{A}_i is a matrix of the respective coefficients in period $t - i$. ε_t is a three-dimensional vector with serially uncorrelated and mutually uncorrelated errors. We assume that matrix \mathbf{A}_0 has a recursive structure such that

¹³Onshore/Offshore

¹⁴Quarterly CPI for the US is taken from OECD statistics.

¹⁵For a discussion on the identification of shocks and estimation of VARs see Kilian (2011)

the reduced form errors \mathbf{e}_t can be decomposed according to $\mathbf{e}_t = \mathbf{A}_0^{-1} \boldsymbol{\varepsilon}_t$:

$$\begin{pmatrix} e_t^{p_o} \\ e_t^D \\ e_t^{cD} \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} = 0 & a_{13} = 0 \\ a_{21} & a_{22} & a_{23} = 0 \\ a_{31} & a_{32} & a_{33} \end{pmatrix} \begin{pmatrix} \varepsilon_t^{p_o} \\ \varepsilon_t^{A_s} \\ \varepsilon_t^{A_d} \end{pmatrix}$$

Note that we provide some intuition on the origin of the structural shocks, in which the variation of the reduced form errors is decomposed, in the theoretical discussion above. In particular, see the representation of demand and supply schedules in equation 6.3 and the discussion below.

The first structural shock, $\varepsilon_t^{p_o}$, is an *oil price demand shock* which is defined by unpredictable innovations to the oil price. We assume that the oil price does not respond contemporaneously (within a quarter) to shocks originating in the upstream sector, shocks to drilling activity or cost of drilling. This assumption is justified by the fact that increased activity in drilling exploration wells does not immediately translates into a change in oil and gas supply. It takes much longer than a quarter, and typically more than a year, before production starts after exploring and assessing new reserves. Alternatively, one could argue that expectations of market participants about new discoveries effect the oil price. But it usually takes more than a quarter before a reasonable estimate of the size of an oil field is formed, and it takes even longer before total drilling costs can be determined (Adelman 1990). Formally, we assume a_{12} and a_{13} to be zero.

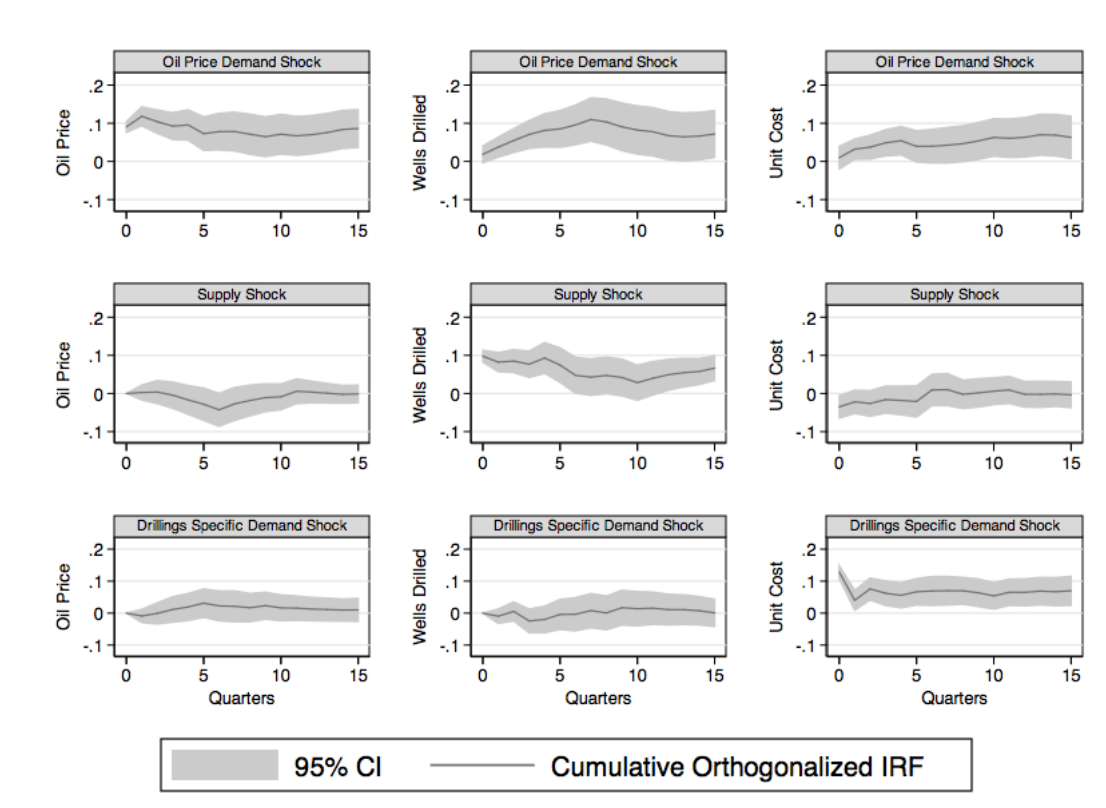
The remaining variation in the reduced form errors - after controlling for the variation in the oil price - is decomposed into $\varepsilon_t^{A_s}$, a *supply shock*, and $\varepsilon_t^{A_d}$, a *drilling specific demand shock*. We think of a supply shock being driven by technological shocks, e.g. advances in deep sea offshore drilling or development of horizontal drilling. The latter is referred to as a drilling specific demand shock to emphasize that the shock originates in the upstream sector and is not driven by fluctuations in the oil price. We think of a drilling specific demand shock as being mainly driven by regulatory changes in the upstream sector: e.g. increase in offshore drilling following the expansion of area available for leases, the nationalization of an oil and

gas industry restricting access to specific regions, an increase in taxes or tougher safety regulations. To disentangle the two we assume that the supply curve for exploration drillings is vertical in the short run (within a quarter). Hence, only supply shocks can affect the number of wells drilled and costs of drilling contemporaneously. More formally, we assume a_{23} to be zero. This assumption requires a discussion.

Obviously, overall supply of engineers and rigs on the global level is limited in the short run: most certainly it requires more than a quarter to educate an additional engineer (Financial Times 2014); from industry specialists we know that it typically requires more than a quarter to build a small rig and up to 3 years to build an offshore rig. One could still argue that supply can adjust at the intensive or the extensive margin in the presence of spare capacities. Kellogg (2011) reports that the scope for adjustments at the intensive margin is rather limited as hired drilling crews typically operate 24 hours a day and 7 days a week, rotating crews in 8 hours shifts. It is more difficult to argue that there are no spare capacities which would allow for an adjustment at the extensive margin. (Ishii 2011) studies the oil and gas industry in the US and reports the existence of spare capacities which are, however, costly to activate. This suggests the existence of a steep but not vertical supply curve of drilling services. However, he focuses on all types of drilling rigs, including rigs used to drill production wells, in well explored oil and gas fields in the US. We, on the other hand, focus on the subsample of exploration wells for all the countries outside the US. Thus, we argue that oil companies exploring new territories outside well developed oil and gas fields and the existence of capacity constraints in most developing countries (e.g. Russia, Central Asia, Africa, etc.) will most certainly face vertical supply curves in the short run (within a quarter). Mainly because rigs have to be utilized, or even constructed, transported to a new region, and a new crew has to be hired and possibly educated to operate the rig in the new territory. Our interviews with energy specialists suggest that this process requires more than a quarter, and usually up to 3 quarters.

Before estimating the VAR model we conducted the usual specification tests on the stability of the VAR, the normality and the serial correlation of the errors and the optimal number of lags. In order to achieve normality and no serial correlation of the variables, we introduced

Figure 6.4: Responses to One-Standard-Deviation Structural Shocks



one dummy variable to capture the negative oil price shock in the last quarter of 2008. After introducing the dummy we cannot reject the hypothesis of all the errors being normally distributed and serially uncorrelated up to 8 lags at the 5% significance level. We use more lags than suggested by standard tests on lag selection (only according to the LR test is it optimal to introduce 6 lags). This is because we expect that it requires up to 6 quarters until the activity of oil and gas firms fully adjusts to a shock in the oil price. Finally, instead of relying on asymptotic theory we bootstrap the standard errors.

6.6 Results

The impulse response functions of our baseline results are presented in Figure 6.4. Following a structural *oil price demand shock*, $\varepsilon_t^{p_o}$, the number of wells drilled and average cost of drilling a metre significantly increase from the second quarter onwards. Also, note that the increase

is permanent in both variables. Our estimates suggest a long run price elasticity of supply of approximately 1-2. Following a 10% increase in the oil price, average cost of drilling a metre and the number of wells drilled increases by 6% and 8%, respectively.

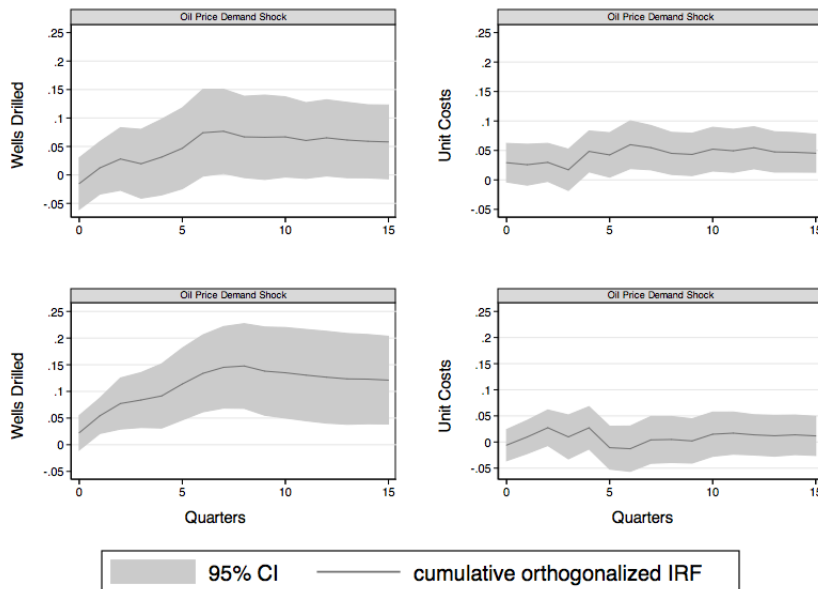
Following a structural *supply shock*, $\varepsilon_t^{A_s}$, costs of drilling decreases. However, the effect on costs per metre drilled remains rather small and predominantly insignificant, suggesting that the demand curve is relatively flat. Note that the effect on the oil price is, as expected, negative and significant. However, the effect remains transitory and turns insignificant after approximately 2 years.

Following a structural *drilling specific demand shock*, $\varepsilon_t^{A_d}$, the effect on the number of wells drilled and the real price of oil is hump-shaped but insignificant. Most importantly, the effect on the real price of oil does not appear to be permanent.

Note that the response of the number of wells drilled to an oil price demand shock and a drilling specific demand shock differs, implying a different price elasticity of supply. This difference is not captured by our theoretical framework which is too simple to reflect every nuance of the oil and gas industry. Thus, this result merits some discussion. There are at least three reasons why the responses differ. First, an oil and gas company typically does not react instantaneously to an innovation in the price of oil. Mainly because of the volatility of the oil price which taught oil and gas companies to be cautious in interpreting changes in the oil price as being persistent. The hesitation allows firms on the supply side of the upstream sector to prepare for the boom and adjust for the upcoming increase in demand for drillings (Ishii 2011). Thus, the delay in making investment decisions allows firms on the supply side to flatten the supply curve.

Second, our interviews with industry specialists suggest that oil companies operate under relatively tight budget constraints. Essentially, this is driven by the well known link between debt ratios (e.g. total debt/total assets) and interest rates paid on capital employed in the production process (see for example (Merton 1974)). Because the activity of oil companies is

Figure 6.5: Responses to One-Standard-Deviation Structural Oil Price Shock: **Top row** represents the response in lower income countries (GDP per capita < 3000). **Bottom row** represents the response in upper income countries (GDP per capita > 12000)



traditionally capital intensive and exploration is uncertain in its results this link is particularly important in the oil and gas industry (Adelman 1962). A fact investors are very well aware of (Dumont 2014). Thus, an increased demand for drillings in new territories following a drilling specific demand shock (which is not associated with a relaxation of the budget constraint as opposed to an oil price driven demand shock) is typically matched by a decrease in drilling activities in other regions. In fact, this discussion suggests that activity should not be significantly affected by a drilling specific demand shock in the long run, and not only in the short run.

Third, the drilling specific demand shock is inter alia designed to capture regulatory changes: the nationalization of an oil and gas industry, the opening up of a country to international oil companies, the expansion of areas available for leases, etc. We have argued that following regulatory changes oil and gas companies face a vertical supply curve, usually for much longer than 3 months. On the other hand, following an unanticipated increase in the oil price, oil and gas companies typically get involved in already scheduled projects

which have not been approved in the past due to capital constraints. Thus, we argue that drilling specific demand shocks and oil price demand shocks result in a different combination of projects which in turn face different supply curves.

One way to observe the heterogeneity in the slope of the supply curves in the upstream sector is the decomposition of our sample. We decompose our sample into drillings which took place only in upper income countries or only in lower income countries and repeat the above exercise. The response of drilling activity and costs to a structural oil price demand shock is presented in Figure 6.5 (top row for lower income countries and bottom row for upper income countries). The response in activity is similar across groups. The same does not apply to the response in costs of drilling. The effect on costs in upper income countries is not significantly different from zero. On the other hand, the increased activity in lower income countries is accompanied by a permanent increase in drilling costs. This could be due to the existence of supply constraints in lower income countries. This may also partly explain why, despite a similar trend in activity (see Figure 6.9 in the Appendix), drilling costs have accelerated recently in developing countries but remained nearly constant in upper income countries (see Figure 6.10 in the Appendix).

6.7 Conclusion

We use micro data from Wood MacKenzie to construct two quarterly time series capturing the total number of exploration wells drilled by the 25 major oil and gas companies and (ii) the average cost of drilling a metre. In combination with the real price of oil we estimate a three dimensional structural VAR model. The proposed model allows us to decompose the variation in the reduced form errors from estimating the VAR into three structural shocks by assuming a recursive structure: an *oil price demand shock*, a *supply shock* and a *drilling specific demand shock*. We estimate the dynamic effects of these shocks on drilling activity, costs of drilling and the real price of oil.

We have three main results. First, our results suggest an upwards sloping supply curve in the upstream sector of the oil and gas industry with a long run price elasticity of supply of 1-2. We also observe the existence of significant heterogeneity in the slope of the supply curve. In particular, our results seem to be driven by less developed countries, which have a steeper supply curve. Second, the demand curve in the upstream sector appears to be flat. Third, supply shocks and drilling specific demand shocks do *not* permanently affect the oil price. This is particularly interesting because it calls into question the theoretically expected link between extraction costs and the price of oil (Krautkraemer 1998). We follow Kilian (2009) in arguing that this result may be explained by the endogenous expansion of crude oil production as a reaction to an increase in drilling costs.

6.8 Appendix

Figure 6.6: Number of wells drilled

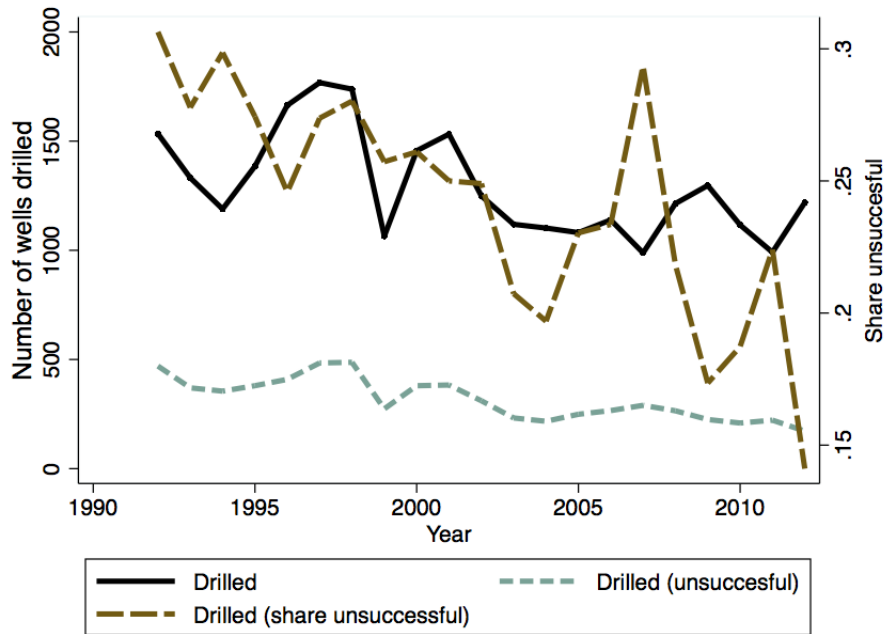
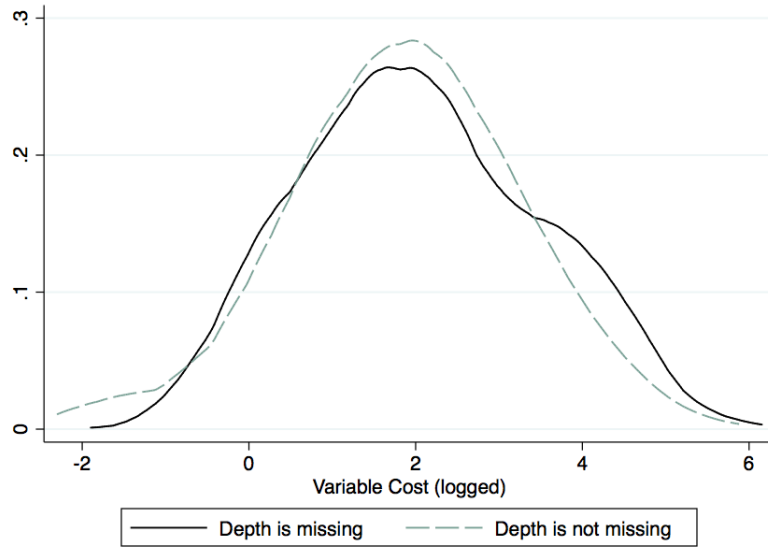


Table 6.2: Companies in the sample

Company	State Owned	Country	Year of Foundation
Anadarko	N	US	1959 (back to 1928)
Aramco	Y	Saudi Arabia	1933
BP	N	UK	1909
Chevron	N	US	1984 (back to 1876)
ConocoPhillips	N	US	2002 (back to 1875)
ENI	N	Italy	1953
ExxonMobile	N	US	1999 (back to 1870)
Gazprom	Y	Russia	1989 (back to 1965)
NIOC	Y	Iran	1948
KOC	Y	Kuwait	1934
Lukoil	Y	Russia	1991
NNPC	Y	Nigeria	1977
PDVSA	Y	Venezuela	1976
Pemex	Y	Mexico	1938
PetroChina	Y	China	1999 (back to 1988)
Petrobras	Y	Brazil	1953
Petronas	Y	Malaysia	1974
QP	Y	Qatar	1974
Rosneft	Y	Russia	1993
Shell	N	Netherlands	1907
Sinopec	Y	China	2000
Sonatrach	Y	Algeria	1963
Statoil	N	Norway	1972
Total	N	Total	1924

Figure 6.7: PDF of variable costs (Million \$US, logged)



dis

Figure 6.8: PDF of total depth (metres, logged)

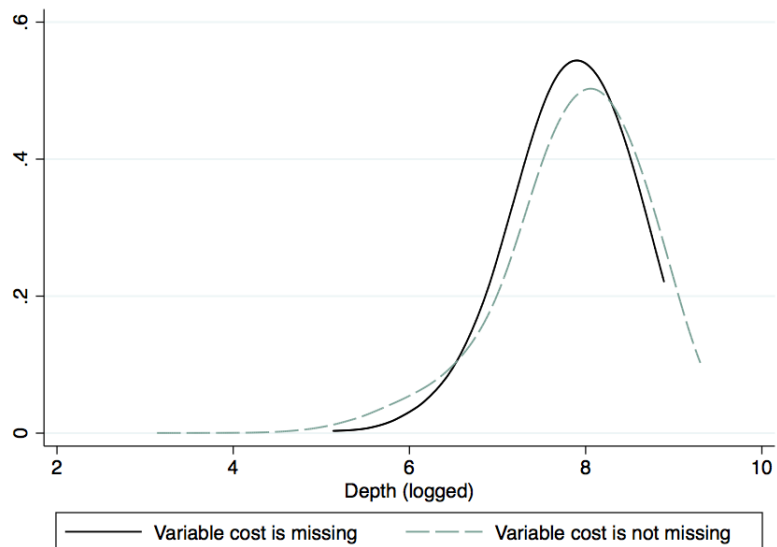


Figure 6.9: Number of wells drilled by level of development

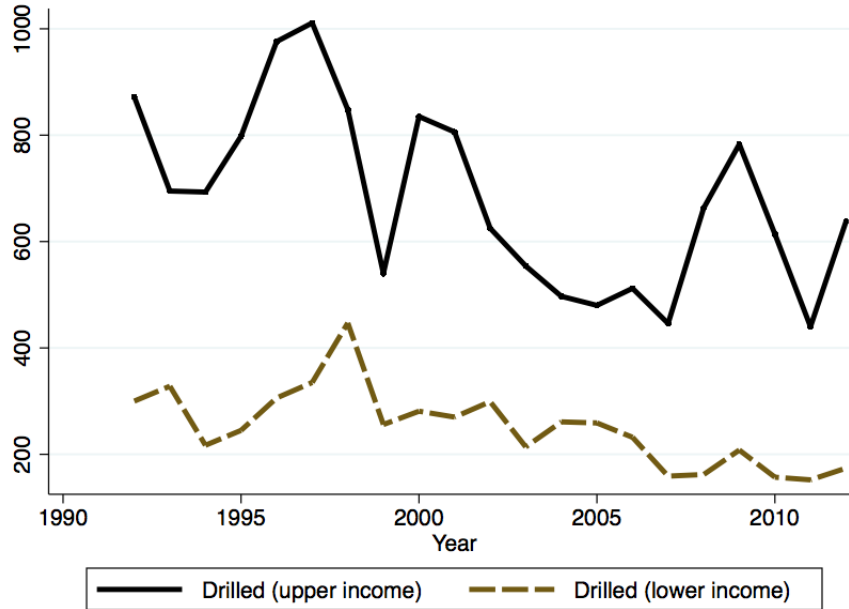
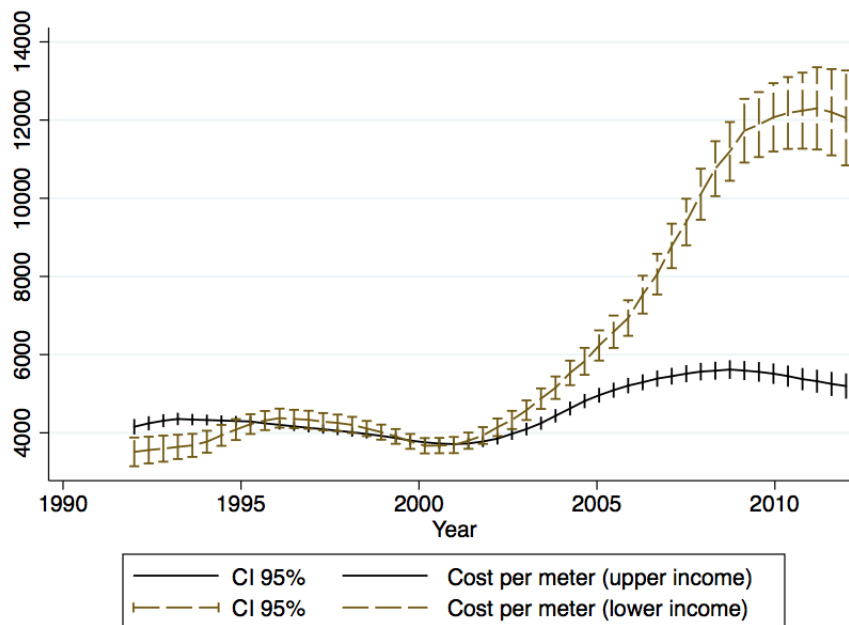


Figure 6.10: Costs of wells drilled by level of development



Bibliography

- Abeler, J., A. Falk, L. Goette, and D. Huffman (2011). Reference points and effort provision. *The American Economic Review* 101(2), 470–492.
- Adelman, M. A. (1962). *The Economics of Petroleum Supply: Papers by MA Adelman, 1962-1993*. MIT press.
- Adelman, M. A. (1990). Mineral depletion, with special reference to petroleum. *The Review of Economics and Statistics* 72(1), 1–10.
- Ai, C. and E. C. Norton (2003). Interaction terms in logit and probit models. *Economics letters* 80(1), 123–129.
- Allcott, H. and D. Keniston (2013). Dutch disease or agglomeration? the local economic effects of natural resource booms in modern america. Technical report, Working Paper.
- Anderson, S. T., R. Kellogg, and S. W. Salant (2014, March). Hotelling under pressure. Working Paper.
- Angrist, J. and J. Pischke (2009). *Mostly Harmless Econometrics*. Princeton University Press.
- Aragón, F. M. and J. P. Rud (2013). Natural resources and local communities: evidence from a peruvian gold mine. *American Economic Journal: Economic Policy* 5(2), 1–25.
- Atal, V., K. Basu, J. Gray, and T. Lee (2010). Literacy traps: Society-wide education and individual skill premia. *International Journal of Economic Theory* 6(1), 137–148.
- Atkin, D. G. (2012). Endogenous skill acquisition and export manufacturing in mexico. Technical report, National Bureau of Economic Research.
- Auty, R. (2008). Improving the beneficial socio-economic impact of hydrocarbon extraction on local/regional development in caspian economies. In B. Najman, R. Pomfret, and

- G.Raballand (Eds.), *The Economics and Politics of Oil in the Caspian Basin*. Routledge.
- Auty, R. M. (2001). *Resource abundance and economic development*. Oxford University Press.
- Beatty, C. and S. Fothergill (1996). Labour market adjustment in areas of chronic industrial decline: the case of the uk coalfields. *Regional studies* 30(7), 627–640.
- Beatty, C., S. Fothergill, and R. Powell (2007). Twenty years on: has the economy of the uk coalfields recovered? *Environment and Planning A* 39(7), 1654.
- Betts, J. R. and L. L. McFarland (1995). Safe port in a storm: The impact of labor market conditions on community college enrollments. *Journal of Human Resources*, 741–765.
- Beynon, Huw, H. E. and R. Hudson (1999). Working together to achieve more: Strategies for facilitating regeneration. <http://www.cardiff.ac.uk/socsi/whoswho/beynon-coalfieldswkgpapers.html>. Accessed: 2014-08-03.
- Birdsall, N., T. C. Pinckney, and R. Sabot (2000). *Natural resources, human capital, and growth*. Carnegie Endowment for International Peace.
- Black, D., T. McKinnish, and S. Sanders (2005a, April). The economic impact of the coal boom and bust. *The Economic Journal* 115, 449–476.
- Black, D. A., T. G. McKinnish, and S. G. Sanders (2005b). Tight labor markets and the demand for education: Evidence from the coal boom and bust. *Indus. & Lab. Rel. Rev.* 59, 3.
- Blanchard, O. J. and L. F. Katz (1992). Regional evolutions. *Brooking's papers on economic activity* 1(1992), 1–61.
- Blanco, L. and R. Grier (2012). Natural resource dependence and the accumulation of physical and human capital in latin america. *Resources Policy* 37(3), 281–295.
- Bobonis, G. J. and F. Finan (2009). Neighborhood peer effects in secondary school enrollment decisions. *The Review of Economics and Statistics* 91(4), 695–716.
- Braakmann, N. (2011, September). Oil discoveries and economic development in the us, 1880 to 1910. Working Paper.
- Broadberry, S. and T. Leunig (2013). The impact of government policies on uk manufacturing since 1945. <https://www.gov.uk/government/uploads/system/uploads/>

attachment_data/file/277158/ep2-government-policy-since-1945.pdf.

Accessed: 2014-08-03.

- Brunnschweiler, C. N. and E. H. Bulte (2008). The resource curse revisited and revised: A tale of paradoxes and red herrings. *Journal of Environmental Economics and Management* 55(3), 248–264.
- Cabrales, A. and E. Hauk (2011). The quality of political institutions and the curse of natural resources*. *The Economic Journal* 121(551), 58–88.
- Caselli, F. and G. Michaels (2013). Do oil windfalls improve living standards? evidence from brazil. *American Economic Journal: Applied Economics* 5(1), 208–38.
- Champion, A. G. and A. R. Townsend (1990). *Contemporary Britain: a geographical perspective*. Edward Arnold London.
- CNPC (2012, April). Cnpc in kazakhstan. <http://www.cnpc.com.cn/eng/cnpcworldwide/euroasia/Kazakhstan/>.
- Collier, P. (2013, April). Under pressure. <http://www.imf.org>.
- Collier, P. and A. Hoeffler (2004). Greed and grievance in civil war. *Oxford Economic Papers* 56(4), 563–595.
- Conley, T. G. (1999). Gmm estimation with cross sectional dependence. *Journal of econometrics* 92(1), 1–45.
- Cook, C. and J. Stevenson (2001). *The Longman handbook of modern British history, 1714-2001*. Pearson Education.
- Cordon, W. M. and J. P. Neary (1982, December). Booming sector and deindustrialisation in a small open economy. *The Economic Journal* 92(368), 825–848.
- CPC (2011, October). Caspian pipeline consortium. <http://www.cpc.ru/>.
- Crawford, V. P. and J. Meng (2011). New york city cab drivers' labor supply revisited: Reference-dependent preferences with rational expectations targets for hours and income. *The American Economic Review* 101(5), 1912–1932.
- Davis, G. A. (1995). Learning to love the dutch disease: Evidence from the mineral economies. *World development* 23(10), 1765–1779.

- Deaton, A. (1992). Household saving in ldc's: Credit markets, insurance and welfare. *The Scandinavian Journal of Economics*, 253–273.
- Dellavigna, S. (2009). Psychology and economics: Evidence from the field. *Journal of Economic Literature* 47(2), 315–372.
- Di Pietro, G. (2006). Regional labour market conditions and university dropout rates: Evidence from Italy. *Regional Studies* 40(6), 617–630.
- Di Tella, R. and R. MacCulloch (2006). Some uses of happiness data in economics. *The Journal of Economic Perspectives* 20(1), 25–46.
- Diener, E. (1984). Subjective well being. *Psychological Bulletin* 95(3), 542–575.
- Duesenberry, J. (1949). *Income, Savings and the Theory of Consumer Behaviour*. Harvard University Press.
- Dumont, C. (2014). 4 leverage ratios used in evaluating energy firms. http://www.investopedia.com/articles/fundamental-analysis/12/4-leverage-ratios-used-in-evaluating_energy_firms.asp. Accessed: 2014-08-03.
- Duncan, B. (1965). Dropouts and the unemployed. *The Journal of Political Economy* 73(2), 121–134.
- Easterlin, R. (1974). Does economic growth improve the human lot? some empirical evidence. *Nations and households in economic growth* 89.
- Easterlin, R. (1995). Will raising the incomes of all increase the happiness of all? *Journal of Economic Behavior and Organization* 27, 35–47.
- Easterlin, R. (2001, July). Income and happiness: Towards a unified theory. *The Economic Journal* 111, 465–484.
- EIA (2012, April). Crude oil price. <http://www.eia.gov/>.
- Elborgh-Woytek, M. K., M. M. Newiak, M. K. Kochhar, M. S. Fabrizio, K. Kpodar, M. P. Wingender, M. B. J. Clements, and M. G. Schwartz (2013). *Women, work, and the economy: macroeconomic gains from gender equity*. International Monetary Fund.
- Emerson, P. M. and B. McGough (2011). *Learning about education*. IZA Discussion Paper.

- Emery, H., A. Ferrer, and D. Green (2011, September). Long term consequences of natural resource booms for human capital accumulation. Canadian Labour Market and Skills Researcher Network. Working Paper.
- Esanov, A. and K. Kuralbayeva (2008). Ricardian curse of the resource boom: The case of kazakhstan 2000-2008. OxCarre Research Paper 43.
- Federman, M. and D. I. Levine (2004). The effects of industrialization on education and youth labor in indonesia. *Contributions to Macroeconomics* 5(1).
- Fernihough, A. and K. H. O'Rourke (2014). Coal and the european industrial revolution. Technical report, National Bureau of Economic Research.
- Ferrer-I-Carbonel, A. and P. Frijters (2004). How important is methodology for the estimates of the determinants of happiness? *Economic Journal* 114, 641–659.
- Ferrer-i Carbonell, A. (2005). Income and well-being: an empirical analysis of the comparison income effect. *Journal of Public Economics* 89(5), 997–1019.
- Fields, G. S. (1974). The private demand for education in relation to labour market conditions in less-developed countries. *The Economic Journal* 84(336), 906–925.
- Financial Times (2014, July 16,). Terrifying oil skills shortage delays projects and raises risks. July 16.
- Frankel, J. A. (2010). The natural resource curse: a survey. Technical report, National Bureau of Economic Research.
- Frey, B. S. (2008). *A Revolution in Economics*. SPi Publisher Services.
- Frijters, P, J. Haisken-DeNew, and J. Shields (2004). Money does matter! evidence from increasing real income and life satisfaction in east germany following reunification. *American Economic Review* 94, 730–740.
- Gill, D. and V. Prowse (2012). A structural analysis of disappointment aversion in a real effort competition. *The American Economic Review* 102(1), 469–503.
- Glewwe, P. and H. G. Jacoby (2004). Economic growth and the demand for education: is there a wealth effect? *Journal of Development Economics* 74(1), 33–51.

- Glyn, A. (1988). The economic case against pit closure. In D. J. Cooper and T. Hopper (Eds.), *Debating Coal Closures: Economic Calculation in the Coal Dispute, 1984-5*. Cambridge University Press.
- Glyn, A. and S. Machin (1997). Colliery closures and the decline of the uk coal industry. *British Journal of Industrial Relations* 35(2), 197–214.
- Goderis, B. and S. W. Malone (2011). Natural resource booms and inequality: Theory and evidence. *Scandinavian Journal of Economics* 113(2), 388–417.
- Greenstone, M., R. Hornbeck, and E. Moretti (2010). Identifying agglomeration spillovers: Evidence from winners and losers of large plant openings. *Journal of Political Economy* 118(3), 536–598.
- Günther, I. and A. Launov (2012). Informal employment in developing countries: opportunity or last resort? *Journal of development economics* 97(1), 88–98.
- Guriev, S. and E. Zhuravskaya (2009, September). (un)happiness in transition. *Journal of Economic Perspectives* 23(2), 143–68.
- Gurr, T. (1970). *Why Men Rebel*. Princeton, NJ: Princeton University Press.
- Gylfason, T. (2001). Natural resources, education, and economic development. *European Economic Review* 45(4), 847–859.
- Hall, P. (2007). Evolution of strategic planning and regional development. In H. T. Dimitriou and R. Thompson (Eds.), *Strategic Planning for Regional Development in the UK*. Routledge.
- Hall, R. E. (1999). Labor-market frictions and employment fluctuations. *Handbook of macroeconomics* 1, 1137–1170.
- Hamilton, J. D. (2008). Understanding crude oil prices. Technical report, National Bureau of Economic Research.
- Hargreaves, S. (2013, April). 10 most expensive energy projects in the world. <http://www.money.cnn.com>.
- Harris, J. R. and M. P. Todaro (1970). Migration, unemployment and development: a two-sector analysis. *The American Economic Review* 60(1), 126–142.

- Heath, R. and A. M. Mobarak (2012). Does demand or supply constrain investments in education? evidence from garment sector jobs in bangladesh. Technical report, Working Paper, Department of Economics, University of Washington and School of Management, Yale University.
- Helman, C. (2013). The world's 25 biggest oil companies. *Forbes*.
- Higham, N. (2014). Cabinet papers reveal 'secret coal pits closure plan'. <http://www.bbc.co.uk/news/uk-25549596>. Accessed: 2014-08-03.
- Hirschman, A. O. (1958). *The Strategy of Economic Development*. Yale University Press.
- Hotelling, H. (1931). The economics of exhaustible resources. *Journal of Political Economy* 39.
- Humphreys, M. (2005). Natural resources, conflict, and conflict resolution uncovering the mechanisms. *The Journal of Conflict Resolution* 49(4), 508–537.
- IMF (2003, March). Republic of kazakhstan: Report on observance of standards and codes. IMF Country Report No. 03/78.
- IMF (2005, July). Republic of kazakhstan: Statistical appendix. IMF Country Report No. 05/239.
- IMF (2009, June). Republic of kazakhstan: 2009 article iv consultation—staff report; supplement; and public information notice on the executive board discussion. IMF Country Report No. 09/300.
- Ishii, J. (2011, March). Useful excess capacity? an empirical study of us oil & gas drilling. Working Paper.
- Jacobsen, G. D. and D. P. Parker (2013). The economic aftermath of resource booms: Evidence from boomtowns in the american west. Working Paper.
- Jensen, R. (2010a). The (perceived) returns to education and the demand for schooling. *The Quarterly Journal of Economics* 125(2), 515–548.
- Jensen, R. T. (2010b). Economic opportunities and gender differences in human capital: Experimental evidence for india. Technical report, National Bureau of Economic Research.

- Kahneman, D. and A. B. Krueger (2006). Developments in the measurement of subjective well-being. *The Journal of Economic Perspectives* 20(1), 3–24.
- Kahneman, D. and A. Tversky (1979, March). Prospect theory: An analysis of decision under risk. *Econometrica* 47(2), 263–292.
- Kaiser, M. J. and A. Pulsipher (2007). A review of the oil and gas sector in kazakhstan. *Energy Policy*.
- Kalyuzhnova, Y. (2008). *Economics of the Caspian Oil and Gas Wealth: Companies, Governments, Policies (Euro-Asian Studies)*. Palgrave MacMillan.
- Kalyuzhnova, Y. and U. Kambhampati (2008). The determinants of individual happiness in kazakhstan. *Economic Systems* 32, 285–299.
- Karachaganak (2012, April). Karachaganak. <http://www.kpo.kz/>.
- Karazhanbasmunai (2012, April). Karazhanbasmunai. <http://www.kase.kz/en/emitters/show/karm>.
- KazMunaiGas (2012, April). Kazmunaigas. <http://www.kmg.kz/>.
- Kellogg, R. (2011). Learning by drilling: Interfirm learning and relationship persistence in the texas oilpatch. *The Quarterly Journal of Economics* 126(4), 1961–2004.
- Kilian, L. (2009). Not all oil price shocks are alike: Disentangling demand and supply shocks in the crude oil market. *The American Economic Review*, 1053–1069.
- Kilian, L. (2011). *Structural vector autoregressions*. CEPR.
- Kőszegi, B. and M. Rabin (2006). A model of reference-dependent preferences. *Quarterly journal of economics* 121(4).
- Krautkraemer, J. A. (1998). Nonrenewable resource scarcity. *Journal of Economic Literature*, 2065–2107.
- Kremer, M. and A. Holla (2009). Improving education in the developing world: What have we learned from randomized evaluations? *Annu. Rev. Econ.* 1(1), 513–542.
- Kuzminov, Y. (2004). *Образование в россии: Что мы можем сделать? вопросы образования (1)*. Technical report, Working Paper.
- Lang, M. (2003, February). Caspian sea poses a tough test. www.epmag.com.

- Le Brun, A., S. R. Helper, and D. I. Levine (2011). The effect of industrialization on children's education. the experience of Mexico. *Review of Economics and Institutions* 2(2).
- Løken, K. V. (2010). Family income and children's education: Using the Norwegian oil boom as a natural experiment. *Labour Economics* 17(1), 118–129.
- Luttmer, E. F. (2005). Neighbors as negatives: Relative earnings and well-being. *The Quarterly Journal of Economics* 120(3), 963–1002.
- Marchand, J. (2011). Local labor market impacts of energy boom-bust-boom in western Canada. *Journal of Urban Economics*.
- Mas, A. (2006). Pay, reference points, and police performance. *The Quarterly Journal of Economics* 121(3), 783–821.
- McBride, M. (2010). Money, happiness, and aspirations: An experimental study. *Journal of Economic Behavior and Organization* 74(3), 262–276.
- Merton, R. C. (1974). On the pricing of corporate debt: The risk structure of interest rates*. *The Journal of Finance* 29(2), 449–470.
- Meruert, M. (2003). *Developing New Rules in the Old Environment*, Chapter 8. Local Government and Public Services.
- MFRK (2011, September). Ministry of finance of the republic of Kazakhstan. <http://www.minfin.kz/>.
- Michaels, G. (2011). The long term consequences of resource-based specialization*. *The Economic Journal* 121(551), 31–57.
- Milanovic, B. (2001). Determinants of cross country income inequality: An augmented kuznets hypothesis. World Bank Policy Research WP
- Ministry of Education of Kazakhstan (2011, February). *Natsional'nyi doklad o sostoyanii i razvitiia obrazovaniya respublikii kazakhstan*. Report.
- Moore, B. and J. Rhodes (1973). Evaluating the effects of British regional economic policy. *The Economic Journal*, 87–110.
- Moore, B., J. Rhodes, and P. Tyler (1986). *The impact of government policies on UK manufacturing since 1945*. HMSO, London.

- Moretti, E. (2010). Local multipliers. *The American Economic Review*, 373–377.
- Moretti, E. (2011). Local labor markets. *Handbook of labor economics* 4, 1237–1313.
- Munayshy Public Foundation (2005). *Petroleum Encyclopedia of Kazakhstan*. Munayshy Public Foundation.
- Munshi, K. and M. Rosenzweig (2006). Traditional institutions meet the modern world: Caste, gender, and schooling choice in a globalizing economy. *The American Economic Review*, 1225–1252.
- Najman, B. et al. (2005). How are oil revenues redistributed in an oil economy? the case of kazakhstan. Working Paper.
- Neumark, D. and W. Wascher (1995). Minimum-wage effects on school and work transitions of teenagers. *The American Economic Review* 85(2), 244–249.
- Nexman, J. and E. Scott (1984). Consistent estimation from partially consistent observations. *Econometrica* 16, 1–32.
- Nguyen, T. (2008). Information, role models and perceived returns to education: Experimental evidence from madagascar. *Unpublished manuscript*.
- NUM (2014). NUM History. <http://www.num.org.uk/page/History-NumHistory>. Accessed: 2014-08-03.
- Ospanova, S. (2012, February). Local content policy: Kazakhstan review. Working Paper.
- Oster, E. and B. M. Steinberg (2013). Do it service centers promote school enrollment? evidence from india. *Journal of Development Economics*.
- Oswald, A. J. (1997). Happiness and economic performance*. *The Economic Journal* 107(445), 1815–1831.
- Panorama (2012, April). Local content requirements. <http://panorama.kz/>.
- Parker, M. and J. Surrey (1993). The october 1992 coal crisis and uk energy policy. *The Political Quarterly* 64(4), 396–409.
- PetroKazakhstan (2012, April). Petrokazakhstan. <http://www.secinfo.com/dsvrh.z12m.d.htm>.
- Pindyck, R. S. (1999). The long-run evolution of energy prices. *The Energy Journal*, 1–27.

- Pineda, J. and F. Rodríguez (2010). *Curse Or Blessing?: Natural Resources and Human Development*. United Nations Development Programme New York.
- Pomfret, R. (2005). Kazakhstan's economy since independence: Does the oil boom offer a second chance for sustainable development? *Europe-Asia Studies* 57(6), pp. 859–876.
- Pomfret, R. (2006). *The Central Asian Economies Since Independence*. Princeton University Press.
- Pritchett, L. (2001). Where has all the education gone? *The World Bank Economic Review* 15(3), 367–391.
- Ray, D. (2006). Aspirations, poverty, and economic change. In A. V. Banerjee, R. Benabou, and D. Mookherjee (Eds.), *Understanding Poverty*, Volume 1. Oxford University Press.
- Rees, D. I. and H. N. Mocan (1997). Labor market conditions and the high school dropout rate: Evidence from new york state. *Economics of Education Review* 16(2), 103–109.
- Richardson, R. and S. Wood (1989). Productivity change in the coal industry and the new industrial relations. *British Journal of Industrial Relations* 27(1), 33–55.
- Rizzo, J. A. and R. J. Zeckhauser (2003). Reference incomes, loss aversion, and physician behavior. *Review of Economics and Statistics* 85(4), 909–922.
- Robinson, M. D. and G. L. Clore (2002). Belief and feeling: evidence for an accessibility model of emotional self-report. *Psychological bulletin* 128(6), 934.
- Ross, M. L. (2007). How mineral-rich states can reduce inequality. In M. Humphreys, J. D. Sachs, and J. E. Stiglitz (Eds.), *Escaping the Resource Curse*. Columbia University Press.
- Ross, M. L. (2012). *The Oil Curse: How Petroleum Wealth Shapes the Development of Nations*. Princeton University Press.
- Rutkowski, J. (2011, February). Promoting formal employment in kazakhstan. Mimeo.
- Sachs, J. D. and A. Warner (1995). Natural resource abundance and economic growth. NBER Working Papers 5398, National Bureau of Economic Research.
- Sange Research Center (2009). Vnutrennyaya migratsiya naseleniya respubliky kazakhstan po itogam perepisi naseleniya 2009 goda. Technical report, Mimeo.

- SARK (2005). *Investment Activity in the Republic of Kazakhstan 2001-2004*. Agenstvo Respubliki Kazakhstan po statistike (Statistical Agency of the Republic of Kazakhstan).
- SARK (2008). *Services in Kazakhstan in 2007*. Agenstvo Respubliki Kazakhstan po statistike (Statistical Agency of the Republic of Kazakhstan).
- SARK (2009). *Economic Activity of the people in Kazakhstan*. Agenstvo Respubliki Kazakhstan po statistike (Statistical Agency of the Republic of Kazakhstan).
- SARK (2011, September). Statistical agency of the republic of kazakhstan. <http://www.stat.kz/Pages/default.aspx>.
- Shubin, D. (2005, April). Kazakhstan. final report. <http://web.warwick.ac.uk/russia/Intas/index.htm>.
- Snyder, R. (2001). Scaling down: The subnational comparative method. *Studies in Comparative International Development (SCID)* 36(1), 93–110.
- Solon, G., S. J. Haider, and J. Wooldridge (2013). What are we weighting for? Technical report, National Bureau of Economic Research.
- Stijns, J.-P. (2006). Natural resource abundance and human capital accumulation. *World Development* 34(6), 1060–1083.
- Stijns, J.-P. (2009). Mineral wealth and human capital accumulation: a nonparametric approach. *Applied Economics* 41(23), 2925–2941.
- Surrey, J. (1992). Technical change and productivity growth in the british coal industry, 1974–1990. *Technovation* 12(1), 15–39.
- Suslova, E. and N. Volchkova (2013). Human capital, industrial growth and resource curse. *Industrial Growth and Resource Curse (October 1, 2012)*.
- Teal, F. (2000). *Employment and Unemployment in Sub-Saharan Africa: An Overview*. University of Oxford.
- Teal, F. (2011). Higher education and economic development in africa: A review of channels and interactions. *Journal of African Economies* 20(suppl 3), iii50–iii79.
- Thatcher, M. (1993). *The Downing Street Years*. HarperCollins.
- Tocqueville, A. d. (1986). *L'Ancient Regime*. Robert Laffont.

- Toews, G. (2014, February). Distributional effects of natural resource booms: Evidence from kazakhstan. Working Paper.
- Toews, G. and A. Libman (2014, February). Getting the incentives right: Human capital investment and natural resource booms. Working Paper.
- Toews, G. and A. Naumov (2014). Modeling the upstream sector of the oil and gas industry. Working Paper.
- UNDP (2004). *Human Development Report for Kazakhstan. Education for All: The Key Goal for a New Millenium*. UNDP.
- Van der Ploeg, F. (2011). Natural resources: curse or blessing? *Journal of Economic Literature*, 366–420.
- van der Ploeg, F. and A. Venables (2012, March). Natural resource wealth: The challenge of managing a windfall. *The Annual Review of Economics* 4, 315–37.
- Veblen, T. (1899). *The Theory of Leisure Class*. Modern Library.
- Venables, A. J. (2011). Depletion and development: Natural resource supply with endogenous field opening. Technical report, CESifo working paper: Resource and Environment Economics.
- Vicente, P. (2010). Does oil corrupt? evidence from a natural experiment in west africa. *Journal of Development Economics* 92(1), 28–38.
- Walker, A. (2012). An empirical analysis of resource curse channels in the appalachian region. Available at SSRN 2165375.
- World Bank (2004, June). Dimensions of poverty in kazakhstan, volume ii: Profile of living standards. Report No. 30294-KZ.
- World Bank (2005a, February). Getting competitive, staying competitive: The challenge of managing kazakhstan’s oil boom. Country Economic Memorandum.
- World Bank (2005b, October). Kazakhstan reforming intergovernmental fiscal relations: Lessons from international experience. Report No. 33709-KZ.
- Wren, C. (1996). *Industrial Subsidies: The UK Experience*. Basingstoke: Macmillan.

Wren, C. and J. Taylor (1999). Industrial restructuring and regional policy. *Oxford Economic Papers* 51(3), 487–516.

Yergin, D. (2011). *The prize: The epic quest for oil, money & power*. Simon and Schuster.

Yessenova, S. (2008). Tengiz crude. In B. Najman, R. Pomfret, and G. Raballand (Eds.), *The Economics and Politics of Oil in the Caspian Basin*. Routledge.