



# INEQUALITY, INFORMALITY AND RESOURCE BOOMS: EVIDENCE FROM KAZAKHSTAN

Peter Howie<sup>1</sup> & Zauresh Atakhanova<sup>2</sup>

1. Graduate School of Public Policy, Nazarbayev University
2. School of Mining and Geosciences, Nazarbayev University

28 October 2020

1

---

## Motivation

1. There are limited number of empirical papers that specifically look at the impact of resource abundance on inequality (Farzanegan, and Krieger 2019)
2. Within this field of research, we are interested in transmission channels by which resource booms cause changes in inequality. The empirical literature on this topic is very thin.

2

---

## Our Objectives

- To study the effect of resource abundance on human development by examining the hypothesis

**“Resource booms decrease the inequality of income and consumption by increasing demand for formal sector goods (and employment) within resource-dependent, low- and middle-income countries”**

---

## Our Contribution

We are interested in assessing:

1. The effects of market on the dynamics of informal employment and ...
2. in turn inequality during resource booms.

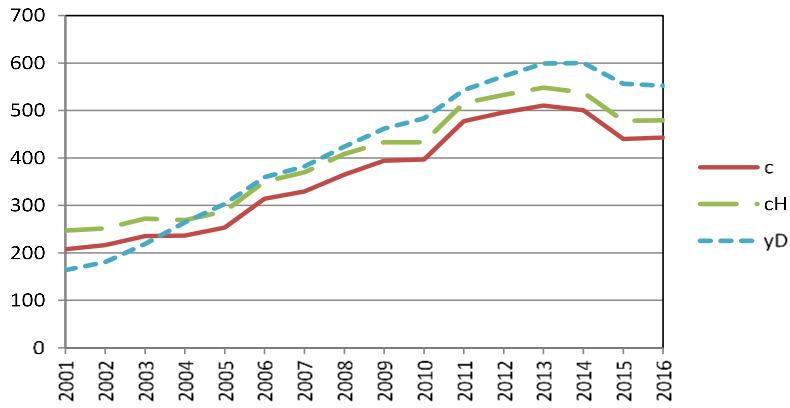
## Understanding these factors is important:

1. To provide governments evidence that the outcomes of policies focusing on formalizing the workforce during a resource boom may suffer from ***spurious correlation***.

## Kazakhstan's Oil Dependence (Sources: IMF, EIA)

	Real GDP	Real oil GDP	Real non-oil GDP	General government revenues	General government oil revenues	Exports of goods and services	Exports of oil and gas condensate	Oil and gas condensate production	Oil price Europe Brent
Unit	% change	% change	% change	% of GDP	% of GDP	bln US dollars	bln of US dollars	mln barrels per day	US dollars per barrel
2002	9.8	22.4	7.6	22.5	4.4	11.6	5	0.97	24.99
2003	9.3	11.6	8.9	25.4	6	14.9	7	1.06	28.85
2004	9.6	15.4	8.4	24.6	7	22.6	11.4	1.22	38.26
2005	9.4	2.3	11	28.6	10.8	30.6	17.4	1.27	54.57
2006	10.7	9.9	10.8	27.5	10.2	41.6	23.6	1.34	65.16
2007	8.9	6.3	8.5	28.8	9.4	51.9	28.1	1.39	72.44
2008	3.2	2.5	3.3	27.8	12.4	76.4	41.5	1.45	96.94
2009	1.2	7.1	0.5	22.1	9.2	48.2	26.2	1.56	61.74
2010	7.3	7.3	7.2	23.9	11.7	65.8	37	1.67	79.61
2011	7.5	1.4	10.2	27.7	14.8	89.5	55.2	1.64	111.26
2012	4.8	-2.2	8	26.9	13.4	91.8	56.4	1.61	111.63
2013	6	2.9	7.2	23.3	11.5	90.9	57.2	1.66	108.56
2014	4.2	0.3	6.3	23.7	11.3	86.9	53.6	1.72	98.97
2015	1.2	-2.6	2.5	16.6	6.6	53	26.8	1.75	52.32
2016	1.1	-1.2	1.8	16.1	4.1	43.6	19.3	1.70	43.64

## Median Income and Consumption Values, KZT1000s (2008)



*c* = household non-durable expenditures; *cH* = *c* + consumption of home-grown food; *yD* = household disposable income

Source: Authors' calculations of Kazakhstan Household Budget Survey data

## Inequality in Household Income and Consumption, 2001 - 2016

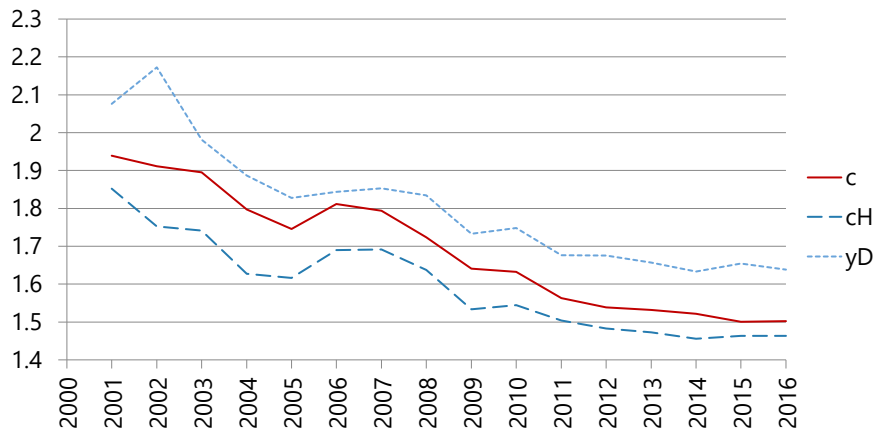
### A. Gini Coefficient



Source: Authors' calculations of Kazakhstan Household Budget Survey data

## Inequality in Household Income and Consumption, 2001 - 2016

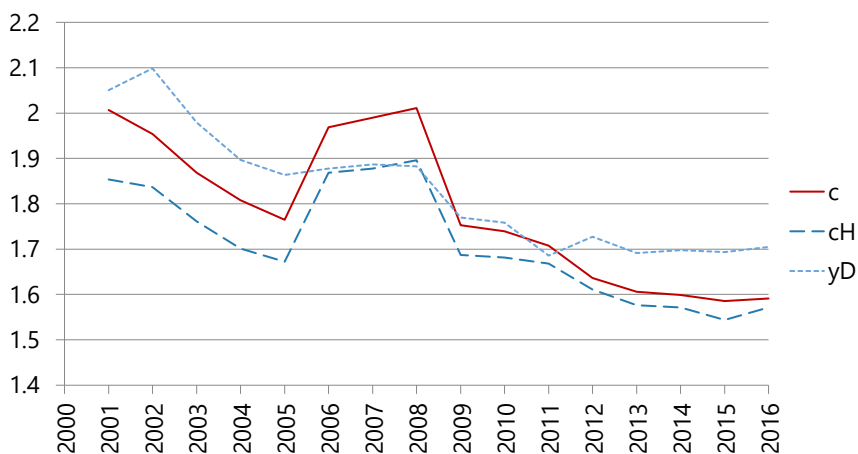
### B. 50/10 Percentile Ratio



Source: Authors' calculations of Kazakhstan Household Budget Survey data

## Inequality in Household Income and Consumption, 2001 - 2016

### C. 90/50 Percentile Ratio



Source: Authors' calculations of Kazakhstan Household Budget Survey data

## Informal workforce, 2004 - 2015

1. To focus on working age adults, we include individuals between the age of 16–58 for women and 16–63 for men.
2. Due to data availability over the study period, we restrict our analysis to employment in the respondents' primary employment.
3. Informality in the case of wage employment, is indicated by the lack of an employment contract
4. A worker is considered *self-employed* if she or he is an employer, own-account worker, unpaid family member, or a member of producers' cooperatives.
5. *Self-employment* is considered informal if the enterprise is not registered.

## Informal workforce, 2004 - 2015

Year	Informally Employed (%)	Informally Employed		Non-Agriculture Informally Employed	
		Agriculture (%)	Non-Agriculture (%)	Wage Workers (%)	Self-Employed (%)
2004	27.7	74.4	25.6	90.6	9.4
2005	25.3	74.1	25.9	91.9	8.1
2006	23.7	75.0	25.0	92.4	7.6
2007	25.4	76.1	23.9	91.8	8.2
2008	23.9	74.0	26.0	92.0	8.0
2009	23.7	72.0	28.0	91.7	8.3
2010	23.1	75.1	24.9	92.5	7.5
2011	18.6	70.4	29.6	93.2	6.8
2012	15.6	79.7	20.3	96.3	3.7
2013	14.1	75.0	25.0	95.8	4.2
2014	18.7	86.0	14.0	96.6	3.4
2015	18.7	86.6	13.4	96.7	3.3

Source: Authors' calculations of Kazakhstan Labor Force Survey data

## Theoretical Model - Consumer Choice & Utility Function

1. Following Banerji and Jain (2007) we consider an economy composed of customers with identical preferences that differ in their levels of income and these income differences imply differences in willingness-to-pay for quality.
2. Individuals derive utility from both a composite commodity,  $x_i$ , which we consider as an agricultural good and an urban-produced differentiated good,  $Z_i$ , which we consider provided both by the urban formal and urban informal sector.
3. Individual  $i$ 's utility function is given by  $u_i = \ln x_i + \theta_i(x_i) Z_i$  where where  $\theta$  is the consumer's willingness to pay for higher quality and is a function of the consumer's income,  $y$ , which is equal to the consumer's unearned income,  $m$ , plus wage income,  $w$ .
4. The subutility derived from the differentiated good takes the following Dixit-Stiglitz utility form:  

$$Z = \left( \int_0^n z(\phi)^\rho d\phi \right)^{\frac{1}{\rho}}$$
 where  $z(\phi)$  is consumption of variety  $\phi$ ,  $n$  is the mass of varieties available to consumers, and  $\rho \in (0, 1)$  is a measure of substitutability

## Theoretical Model - Consumer Choice & Budget Constraint

1. Each individual is assumed to inelastically supply one unit of labor and is either low-skilled,  $L$ , or high-skilled,  $H$
2. Individuals receive income from wages and their share of resource income,  $R$ . Non-wage income,  $m$ , is equal to the return  $\omega_R$  on  $R/L$  units of the resource.
3. Therefore, the budget constraint can be written as:  $p_x x^{j,k} + P_z Z^{j,k} = Y^{j,k} = w^{j,k} L^{j,k} + \omega_r R/L$  where the parameter  $j$  refers to the skill level of the individual,  $k$  refers to the sector in which the individual works, and  $w^{j,k}$  refers to the wage received by the representative worker of skill level  $j$  in sector  $k$ .
4. We follow Banerji and Jain (2007) and assume that low-skilled labor earns less in the informal sector than in the formal sector. This arises because of "binding minimum wages and payroll taxes, which are ... easily evaded by informal firms" (Banerji and Jain, 2007, 237).
5. Furthermore, low-skilled informal sector workers receive a premium with regards to low-skilled agricultural workers (Dasgupta, Bhula-or, and Fakthong 2015; Gagnon, Xenogiani, and Xing 2011).
6. No differences in wages for high-skilled workers between sectors (Lehmann and Zaiceva higher 2013).

## Theoretical Model - Production

---

1. We follow Chao and Takayama (1988) and use a four-sector (i.e., agriculture, formal services, informal services, and natural resources) model with high- and low-skilled labor as the only factors of production.
2. Similar to Chao and Takayama (1988), the agricultural good is produced by unskilled labor alone with constant returns-to-scale technology.
3. We assume that the differentiated-goods industry (i.e., formal and informal goods) consists of  $n$  profit maximizing firms operating in a monopolistic market structure (Helpman, 1984)
4. The natural resource is only sold internationally (i.e., no domestic market) and can be produced without any inputs. That is, we assume that the resource sector is an "exogenous gift" (Benjamin et al. 1989; Stokke 2008; Goderis and Malone 2011; and Howie and Atakhanova 2014)

## Theoretical Model - Implications of a Resource Boom

---

1. Our specification enables changes in utility due to macroeconomic shocks.
2. The effects of a resource boom, which may be associated with increases in  $m$  (and  $y$ ), skews the distribution of consumer types  $\theta$  upwards.
3. This moves some consumers (and both low- and high-skilled workers) from the informal sector to the formal sector. As a result, the informal sector contracts during a resource boom, both relative to the formal sector, and in absolute terms
4. Also, during a resource boom, production will shift from the informal sector to the formal sector due to the increased demand for formal goods and some consumers' income may also increase (because the wage component of income increases).
5. Low-skilled workers will migrate from agriculture to informal/formal sector as well as low-skilled workers will migrate from informal to formal sector due to higher wages.
6. High-skilled workers will migrate from informal to formal sector.



## Theoretical Model - Implications of a Resource Boom

7. Furthermore, the migration of both skilled and unskilled labor is bidirectional between the informal and formal sectors; however, we assume no return migration for low-skilled labor originating in the agricultural sector (Hirvonen and Lilleør 2015).
8. As a result of labor migration, inequality will decrease.

## Empirical Analysis: Basic inequality regression

Following the empirical work of Goderis and Malone (2011) we plan to estimate the following dynamic panel data model:

$$\Delta G_{i,t} = \alpha_i + \gamma_1 G_{i,t-1} + \gamma_2 x_{i,t-1} + \gamma_3 \Delta G_{i,t-1} + \gamma_4 \Delta z_{i,t-1} + \varepsilon_{i,t}$$

- where the subscripts  $i = 1, \dots, N$  and  $t = 1, \dots, T$  index the regions and the quarters in the panel, respectively.
- $G_{i,t}$  represents household inequality in region  $i$  in period  $t$ ;  $\alpha_i$  is the region-specific fixed effect;  $x_{i,t-1}$  is a  $k \times 1$  vector of  $k$  variables that affect inequality in levels;  $z_{i,t-1}$  is a  $m \times 1$  vector of  $m$  variables that affect inequality in differences
- The urban data set includes quarterly values for all 14 regions and the cities of Nur-Sultan and Almaty for the period 2005.Q1 to 2015.Q4. The rural data set includes values for all 14 regions and quarters for which data are available – from 2006.Q1 to 2008.Q4 some oblasts were sampled only three quarters in each calendar year.

## Empirical Analysis: Basic inequality regression

- Our measure of inequality  $G_{i,t}$  is computed either as the Gini coefficient, the p50/p10 ratio, or the p90/p50 ratio and were constructed by using household non-durable and home-produced consumption.
- The  $x_{i,t-1}$  vector includes (1) size of the non-agricultural informal sector; (2) mean education level as a measure of average educational attainment and can proxy the size of the middle class; (3) GRP per capita (in logs) as a proxy for overall development; (5) share of GRP in mining and oil to control type of economic activity; (6) average oil price in the current quarter; and (7) and lagged inequality.
- The  $z_{i,t-1}$  vector includes (1) lagged inequality and (2) per capita health expenditures as a proxy of institutional quality.
- We plan to use Arellano and Bond Systems GMM to estimate the equation.

## Empirical Analysis: The transmission channel

Following Winkelried (2005) we plan to estimate the following dynamic panel data model:

$$INF_{i,t} = \phi_i + \theta_1 INF_{i,t-1} + \theta_2 v_{i,t-1} + \eta_{i,t}$$

- where the subscripts  $i = 1, \dots, N$  and  $t = 1, \dots, T$  index the regions and the quarters in the panel, respectively.
- $INF_{i,t}$  share of non-agricultural informal firms in region  $i$  in period  $t$ ;  $\phi_i$  is the region-specific fixed effect;  $v_{i,t-1}$  is a  $j \times 1$  vector of  $j$  variables that affect informality
- The urban data set includes quarterly values for all 14 regions and the cities of Nur-Sultan and Almaty for the period 2005.Q1 to 2015.Q4. The rural data set includes values for all 14 regions and quarters for which data are available – from 2006.Q1 to 2008.Q4 some oblasts were sampled only three quarters in each calendar year.

## Empirical Analysis: The transmission channel

---

- The  $v_{i,t-1}$  vector includes 1) lagged informal sector size; (2) mean education level as a measure of average educational attainment; (3) GRP per capita (in logs) as a proxy for overall development; (5) share of GRP in mining and oil to control type of economic activity; (6) share of GRP in services; (6) average oil price in the current quarter; and (7) an interaction term of average oil price and a dummy variable that takes a unit value for an increase in oil prices between the period  $t$  and  $t-1$ .
- The  $z_{i,t-1}$  vector includes (1) lagged inequality and (2) per capita health expenditures as a proxy of institutional quality.
- We plan to use Arellano and Bond Systems GMM to estimate the equation.

## What still needs to be done

---

1. Solve the theoretical model
2. Prepare the data for the regressions
3. Run both regressions

# Questions?