Institut für Ost- und Südosteuropaforschung
Institute for East and Southeast European Studies

Arbeitsbereich Ökonomie

IOS Working Papers

No. 329 April 2013

Do Natural Resources Define Convergence Clubs? Empirical Evidence from the Kazakh Regions

Miriam Frey*, Carmen Wieslhuber†, and Daniel Frey‡

*Institute for East and Southeast European Studies and University of Regensburg; Corresponding author's email: frey@ios-regensburg.de

†University of Regensburg; Email: carmen.wieslhuber@ur.de

‡University of Hagen and Bayerische Beamten Lebensversicherung a.G.; Email: daniel.frey@diebayerische.de

Institut für Ost- und Südosteuropaforschung
Institute for East and Southeast European Studies

Landshuter Straße 4
93047 Regensburg

Telefon: (09 41) 943 54-10
Telefax: (09 41) 943 94-27
E-Mail: info@ios-regensburg.de
Internet: www.ios-regensburg.de

Contents

List of Tables

List of Figures

Abstract

This paper deals with the hypothesis that natural resources are important in forming convergence clubs. We check this hypothesis by applying a dependence and an endowment measure of natural resource abundance and a regression tree analysis. The results indicate that for the Kazakh regions natural resources indeed play an important role in forming convergence clubs. It is further shown that rather natural resource endowment than resource dependence determines initial conditions and thereby convergence clubs.

JEL-Classification: O13; O47; O53
Keywords: Convergence clubs; Natural resources; Regression tree; Kazakhstan

The authors thank Barbara Dietz, Ksenija Gatskova, Vladimir Otrachshenko, Achim Schmillen and Manuela Troschke as well as conference and seminar participants in Dublin, Istanbul, Regensburg and Shanghai for helpful comments and suggestions.

1 Introduction

Club convergence is an important issue in the economic growth literature as it is shown in a wide range of empirical studies on convergence that there is a considerable amount of countries and regions which do not converge according to the concepts of σ- and β-convergence.[1,2] This is also the case for the Kazakh regions. As Frey and Wieslhuber (2011) point out there is no σ- and absolute β-convergence across the Kazakh regions.[3,4] Therefore, the focus of this paper is on the analysis of club convergence in Kazakhstan.[5]

The term 'convergence club' was first introduced by Baumol (1986). Roughly spoken, the idea behind the concept of club convergence is that there are multiple steady states[6] a country or region[7] may approach. In which 'basin of attraction' (Galor, 1996) a region is, is determined by the initial conditions at the beginning of the growth path.[8] To understand which initial conditions are potentially important in case of the Kazakh regions, a closer look at their characteristics is needed.

The Kazakh regions are very heterogenous in terms of geography, population and economic activity. The regions in southern Kazakhstan (Almaty, Zhambyl and South Kazakhstan) are industrially underdeveloped and dominated by agricultural activities as Ursulenko (2010) explains. At the same time these regions are the most populated ones in Kazakhstan. Kostanai, North Kazakhstan and Akmola produce the highest gross agricultural output. The regions with a strong industrial sector (East Kazakhstan, Karaganda and Pavlodar) are located in the north and the center of Kazakhstan, whereas the oil- and gas-extracting regions (Aktobe, Atyrau, Kyzylorda, Mangistau and West Kazakhstan) are based in the western part of the country (Roudoi et al., 2011). Natural resources (namely oil, gas and coal) play an important role in Kazakhstan[9,10] and they might also influence a

[1] See for example Canova and Marcet (1999), Lee et al. (1997) and Grier and Tullock (1989).

[2] For an introduction to the concepts of σ- and β-convergence see for instance Barro and Sala-i-Martin (2004).

[3] Conditional β-convergence is not a topic here as we assume that the Kazakh regions share the same structural characteristics and should, therefore, have a common steady state. This assumption seems to be reasonable when looking at regions within a country (Barro and Sala-i-Martin, 2004) and is also made for example by Johnson and Takeyama (2001).

[4] However, Aldashev (2011) finds convergence in wages.

[5] Note that Astana city and Almaty city are not included in the following empirical analysis as cities are generally not resource abundant.

[6] For theoretical models with multiple equilibria see for instance Galor (1996) and Azariadis and Drazen (1990).

[7] Note that in the following we only concentrate on regions, even though the arguments would in general also apply to countries.

[8] Empirically, regional club convergence is established for example in case of the European regions (De Siano and Marcella, 2006a; Fischer and Stirböck, 2006), the US States (Johnson and Takeyama, 2001) and the Italian regions (De Siano and Marcella, 2006b).

[9] See for example Agrawal (2008) and Kutan and Wyzan (2005).

[10] Frey and Wieslhuber (2011) illustrate that growth in oil and gas extraction contributed considerably to real GDP growth in Kazakhstan.

region's growth path. Accordingly many empirical studies[11] conclude that resource-rich and resource-poor countries or regions have different patterns of growth.[12] In addition, natural resources are also said to influence other variables like human development indicators (Bulte et al., 2005) and education (Gylfason, 2001), which in turn reflect the initial conditions at the beginning of the growth path.

Because of this and due to the fact that natural resources are unequally distributed across the Kazakh regions we hypothesize that they play an important role in forming convergence clubs. But how should resource abundance be measured? This question became a key issue in the empirical literature on natural resource abundance and economic development.[13] Bond and Malik (2009) point out that 'empirical findings [are] highly sensitive to the choice of resource measures.' In general, two types of natural resource indicators are distinguished. One possibility is to use reserve or production data to capture a region's resource endowment or wealth. The dependence of a region on natural resources is instead determined by taking relative measures as the share of primary commodity exports in total exports or GDP. Therefore, we apply two different measures of resource abundance to identify resource-rich regions. One is based on production data in physical units and is hence considered as an endowment measure. The second measure refers to the value of natural resource production and its relative importance. Accordingly, this indicator is regarded as a dependence measure.

Looking for club convergence poses the question of how to cluster regions. Different approaches are proposed in the literature to identify potential convergence clubs (Harris, 2011). Among these is the Classification and Regression Tree (CART) analysis (Breiman et al., 1984), which endogenously identifies clubs in order to avoid a selection bias problem. Applying regression tree analysis, our hypothesis is tested by endogenously grouping the Kazakh regions according to the variables identifying differences in initial conditions. The resulting groups are then compared to the exogenously identified groups of resource-rich regions. As will be seen, the initial conditions seem to be rather influenced by natural resource endowment than by resource dependence.

This paper provides two main contributions. It extends the existing literature with respect to the analysis of regional club convergence in Kazakhstan and gives some new insights concerning the importance of different resource abundance measures for empirical work on club convergence.

The paper is organized as follows. The empirical strategy consisting of the classification of regions, the identification of split variables and the regression tree analysis is described in section 2. This is followed by some information on the data in section 3. Section 4 provides the results. The last section concludes.

[11] See for example Chambers and Guo (2009), Papyrakis and Gerlagh (2007) and Papyrakis and Gerlagh (2004).

[12] For the countries in Central Asia this is established by Felipe and Kumar (2010).

[13] See for example Bond and Malik (2009) and Brunnschweiler (2008).

2 Empirical Strategy

Testing our hypothesis that natural resources define convergence clubs consists of the following steps. First the Kazakh regions are classified in resource-rich and resource-poor ones applying different measures of resource abundance. In a second step potential split variables are identified. The resulting variables are then used to endogenously group the sample using regression tree analysis. The outcome of this is afterwards compared to the exogenously grouping of regions in resource-rich and resource-poor ones.

2.1 Exogenous Grouping

Empirical studies analyzing the effects of natural resources on economic growth and development use a wide range of different indicators measuring resource abundance. Most commonly used are data on primary commodity exports as share in gross regional product (GRP) or total exports (Bond and Malik, 2009). Based on these indicators various authors established a negative relationship between natural resource abundance and economic growth. However, using a new measure based on resource endowment, Brunnschweiler (2008) concludes that there is a positive relationship between natural resource abundance and economic growth. As the empirical results seem to be sensitive to the measure of natural resource abundance, we construct two different dummy variables[14] to exogenously group the Kazakh regions into resource-rich and resource-poor ones according to a dependence and an endowment measure.

The Kazakh regions highly depending on natural resources are identified in the following way. For each region the share of oil, gas and coal production in total industrial production is computed, based on a 1998–2000 average[15]. A region is then classified as being resource dependent if its oil, gas and coal production accounts for at least 10%[16] of its total industrial production.

For creating a dummy variable indicating resource endowment[17] we instead apply data on physical units (in thousand tonnes and millions of cubic meters[18]) produced by the resource banking industries in each region as share in total physical units extracted in

[14] Bond and Malik (2009) and Isham et al. (2005) also use dummy variables to indicate resource countries.

[15] This is due to data unavailability. Preferably data for the years 1996–1998 would be used to account for the inital conditions at the beginning of the growth path.

[16] Different threshold values for different resource abundance measures can be found in the literature. However the reasons for choosing a certain one are mostly not explained. For example, the IMF (2012) refers to ‘resource-rich developing countries’ if natural resources comprise at least 20% of total exports. Ahrend (2006) instead defines resource-based economies as those in which natural resources account for more than 40% of total exports and 10% of GDP. Because the share in GDP is the measure most closely related to the one we use, we also apply the 10% threshold.

[17] Resource endowment would best be measured in terms of reserves, especially because those can reasonably be considered to be exogenous. As these data are not available on a regional level, data on produced physical units are used. This leaves us with the question of whether production is also exogenous. As the simple correlation between proven oil reserves and oil production on a country level is 0.7, based on a 1998–2008 average, we also consider production to be exogenous.

[18] Note that the different units of measure are not a problem here as we only use shares.

Kazakhstan.[19] Again, natural resources considered are oil, gas and coal. A region is classified as being well-endowed with natural resources if it contributes a minimum of 10% to the amount of total oil, gas or coal produced in Kazakhstan.

Because one might argue that the results of this exogenous grouping are sensitive to the choice of the threshold value we carry out an additional classification as a robustness check. Therefore, we simply devide the 14 Kazakh regions in those, which are endowed with oil and those which are not.

2.2 Split Variables

To identify possible split variables we follow the methodology of Johnson and Takeyama (2001). In doing so we first sort the sample into an ascending order according to each of the variables listed in Table 5 in the Appendix and divide the sorted sample in two equal parts. Each of the resulting subsamples consists of seven regions. We then estimate the following equation allowing for the intercept and slope of the growth regression to vary across the bottom and the upper part of the sample:

$$\frac{1}{t}log(\frac{y_{it}}{y_{i0}}) = \alpha + \alpha^B D_{ji}^B - (\beta + \beta^B D_{ji}^B)log(y_{i0}) + \epsilon_i \quad (1)$$

The dependant variable is the average annual log growth rate of the gross regional product per capita and subscripts t and i denote the time and the region. The right-hand side of Equation 1 includes the log of the initial GRP per capita and D_{ji}^B = 1 if region i is in the bottom half of the sample when it is sorted according to variable j and zero otherwise. We calculate the Wald test statistics for the hypothesis that the parameters of the regression do not vary across the two halves of the sample ($H_0 : \alpha^B = \beta^B = 0$).

2.3 Endogenous Grouping

In 1984, Breiman et al. published their work on classification and regression trees. This methodology is used to create so-called decision trees for a given dataset. Those decision trees are often applied to classify or respectively regress data.

Decision trees are consisting of a set of questions, which split[20] the dataset step by step into smaller and smaller groups. The aim of CART is to find groups that are as homogenous as possible with respect to the endogenous variable. Therefore, the algorithm checks each observation of each exogenous variable in the dataset in order to find the best split. This is identified by the highest degree of homogeneity in the groups resulting after the split. The method is then repeated until a stopping criterion applies or until the resulting groups are totally homogenous.

[19] Note that the calculated shares are 1996–1998 averages.

[20] It should be noted that only binary splits are possible.

In addition, the following characteristics are of interest. Although the endogenous variable has to be numerical, the exogenous variables can either be numerical or categorial. This is caused by the fact that decisions are always binary ones. As CART is a non-parametric procedure, no *a priori* assumptions about the regression function have to be made. CART is also invariant to monotone transformations of the independent variable like taking the logarithm, leaving the structure of the tree unaffected.

The regression tree is created in the following way. Let Y be the endogenous variable and $X_1, \ldots, X_m$ be the exogenous variables. The first thing to do is to specify the stopping criteria (for instance a minimum amount of observations in each group), which define where the splitting of the groups should be stopped. The algorithm then searches for the best splitting criterion in the whole dataset. Therefore, the so-called impurity measure

$$\Delta(h) = \min_{j,s}\left(\min_{c_1}\sum_{x_i \in R_1(j,s)}(y_i - c_1)^2 + \min_{c_2}\sum_{x_i \in R_2(j,s)}(y_i - c_2)^2\right), \tag{2}$$

is used, where $\hat{c}_1 = ave(y_i|x_i \in R_1(j,s))$ and $\hat{c}_2 = ave(y_i|x_i \in R_2(j,s))$ represent the averages of the realizations of the endogenous variable in the resulting groups R_1 and R_2. Subscript $j = 1, \ldots, m$ stands for the exogenous variable picked at the respective split point and s is the realization of the exogenous variable X_j. The parameter h denotes the splitting level, where $h = 1$ for example indicates the impurity measure of the first split. If the best splitting criterion is detected, this step is repeated for both resulting groups R_1 and R_2 until the chosen stopping criterion occurs. The regression values in the final groups are specified by

$$\hat{y}_g = \frac{1}{n}\sum_{i=1}^{n} y_{ig}, \tag{3}$$

where g indicates the group and n represents the total amount of observations in group g.

3 Data

Per capita personal income (see for instance Johnson and Takeyama (2001), Carlino and Mills (1993)) and real per capita GDP (see for example Meliciani and Peracchi (2006), Durlauf and Johnson (1995)) are both used in the empirical work done on convergence. Due to the lack of data on per capita personal income on a regional level, we use the gross regional product per capita for our analysis. This variable measures factor incomes derived from production within a region. In one of their papers on convergence Barro and Sala-i-Martin (1992) state that the empirical results for personal income are almost equal to those for GRP. The already adjusted data in purchasing power parity (PPP) are drawn from the National Human Development Report (UNDP, 2009). It includes data for the 14 Kazakh regions (*oblasts*) for the period 1990–2008. As in 1997 the capital of Kazakhstan was moved from Almaty to Astana, which subsequently was treated as a distinct region, we have a balanced data set of 14 regions only for the period 1997–2008.

Another peculiarity of the Kazakh regional data is that in 1997 an administrative reform was implemented and some regions were united. Therefore, 1998 is the first year for which data for the regions in today's structure are available. Hence we decided to start our analysis with the year 1998 to avoid additional aggregation or disaggregation biases, which might occur in the data before 1998. The weakness of this data set is the limited number of only 14 regions, which might involve a 'degrees-of-freedom problem' (Desdoigts, 1999) when it comes to a cross section analysis. Due to data unavailability we cannot employ data on a more disaggregated level, like districts.

In addition we use data reflecting the different economic and social conditions in the regions in 1998[21] in order to split the sample and classify clubs endogenously. The data for the years 1996–1999 are taken from the Regional Statistical Yearbook of Kazakhstan (The Agency of Statistics of the Republic of Kazakhstan, 2000). This includes data on demography, production, eduction, health care and general life conditions. The variables were chosen due to their role in the growth and convergence literature and data availability. A detailed list of the additional variables can be found in Table 5 in the Appendix.

For testing our hypothesis that natural resources are important in forming convergence clubs, data on resource abundance are needed. The data on industrial production and physical units produced are also drawn from the Agency of Statistics of the Republic of Kazakhstan.

[21] In some cases the characteristics measured in 1999 or the average of 1997 and 1999 are used if data for 1998 were not available.

4 Results

4.1 Exogenous Grouping

Applying the dependence and the endowment measure to the Kazakh regions gives the results presented in Table 1. The resulting groups are almost identical. Karaganda is the only region that makes the difference. With an average share of 3% of oil, gas and coal production in total industrial production, Karaganda would by no means be considered as a resource-rich region according to the dependence measure. However, if the endowment measure is applied Karaganda would be classified as resource-rich as it contributes on average 25% to the total coal production in Kazakhstan.

Table 1: Resource-rich regions

Dependence measure	Endowment measure
Aktobe	Aktobe
Atyrau	Atyrau
Kyzylorda	Kyzylorda
Mangistau	Mangistau
Pavlodar	Pavlodar
West Kazakhstan	West Kazakhstan
	Karaganda

The example of Karaganda shows that the outcome of a classification in resource-rich and resource-poor regions might be sensitive to the choice of the abundance measure. As this also applies to the threshold value we alternatively group the Kazakh regions into those with and without oil reserves. Because this grouping is based on a simple ‘yes or no’ decision criterion no threshold value is needed. The resulting oil regions are listed in Table 2.

Table 2: Oil regions

Oil reserves
Aktobe
Atyrau
Kyzylorda
Mangistau
West Kazakhstan

4.2 Split Variables

The null hypothesis of the Wald test can be rejected in 12 of the 34 cases. Following Johnson and Takeyama (2001) we interpret this result as an indication for the existence of multiple basins of attraction as indicated by the club convergence hypothesis. Most of the 12 variables listed in Table 3 are known from the empirical growth and convergence literature. In addition, some of them are also identified by Johnson and Takeyama (2001) as potential split variables.

Table 3: Significant split variables

10%-significance level	5%-significance level
BIRTH	INFSCHO
COLSTUDPOP	INFSCHOPOP
DEATH	LIVSPACE
HOSBED	
LANGUAGE	
MURDER	
SAVE	
SUICIDE	
VEHIC	

For the following regression tree analysis only the most significant variables, namely INFSCHO, INFSCHOPOP and LIVSPACE are considered.[22] The variables INFSCHO and INFSCHOPOP indicate the number of infant schools per 1,000 persons and the number of children in infant schools per 1,000 persons respectively. Thus, these variables might be seen as proxies for the state of development of the educational system, which is also closely related to the level of human capital. The variable LIVSPACE represents the average living space per person. We interpret this as an indicator for the overall level of economic and social development.

4.3 Endogenous Grouping

Applying the CART algorithm to our dataset sorts the regions into four groups as shown in Figure 1. The ellipses are nonterminal nodes indicating the splitting criteria. Terminal nodes are represented by the rectangles, which include the number of regions in each group. If the observed value of the chosen split variable for region i is put on the left-hand side of the split value and the resulting inequality holds, the region goes to the left. Accordingly, if the respective inequality is not true, the region goes to the right.

[22] As the variables INFSCHO and INFSCHOPOP are highly correlated only INFSCHO is used for the regression tree analysis.

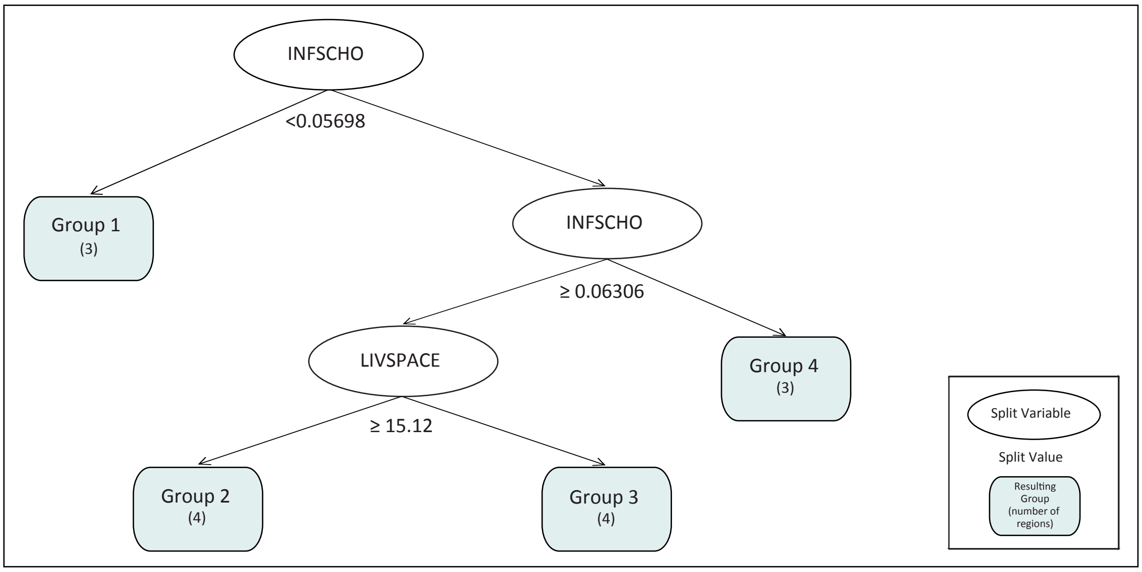

Figure 1: Regression tree

In Figure 1, the first split in the tree is on the variable INFSCHO, which separates the regions into those which had less than 0.05698 infant schools per 1,000 persons in 1998 (left branch of the tree) and those which had more. The second split divides the remaining regions again according to the variable INFSCHO, resulting in a terminal node on the right hand side of the tree (Group 4) and another nonterminal node. For the final split the variable LIVSPACE is picked. Regions, where the average living space in square meters per person was more than 15.12 go for Group 2 and regions with less living space end up in Group 3.

The regression tree shows that using the impurity measure the variable INFSCHO provides the best splitting criterion for the first two splits. This indicates that the variable INFSCHO is more useful in determining different groups of regions than the variable LIVSPACE. The results further suggest that the number of infant schools per 1,000 persons plays an important role in determining the growth path a region is going to tread in the following. This finding is in line with the conclusion of former empirical studies.[23]

Most interestingly the resulting groups (see Table 4) can be characterized in the following way. While the regions in Group 1 and Group 4 are either agricultural or industrial producers, the Group 2 and Group 3 regions are well-endowed with oil, gas or coal (except for North Kazakhstan in Group 2). Group 2 contains the two major producers of coal, Karaganda and Pavlodar, producing together more than 95% of the total physical units extracted in Kazakhstan. All regions in Group 3 are located in the west and southwest of Kazakhstan. In addition, oil is extracted in each of those. Therefore, our results support the findings of Durlauf and Johnson (1995) concerning the ‘geographic homogeneity’ within groups.

[23] See for example Ahrend (2002) and Barro (2001).

Table 4: Resulting groups

Group 1	Group 2	Group 3	Group 4
Almaty	Karaganda	Aktobe	Akmola
South Kazakhstan	North Kazakhstan	Atyrau	East Kazakhstan
Zhambyl	Pavlodar	Kyzylorda	Kostanai
	West Kazakhstan	Mangistau	

As a final step we compare the outcome of the regression tree analysis with the exogenous classification of regions according to a natural resource dependence and an endowment measure. Looking at the groups resulting after the two INFSCHO splits[24] and comparing those two groups (Group 2 and Group 3) to the exogenous classification of regions according to the resource endowment measure shows that the results are almost identical. The only region which would be grouped wrongly when applying the endowment measure would be North Kazakhstan. Applying the dependence measure gives similar results, even though the group of regions well-endowed with natural resources better fits to the outcome of the regression tree analysis. We take this as an indication for the confirmation of our hypothesis and conclude that rather natural resource endowment than dependence is important in forming convergence clubs. This is additionally endorsed by the results of our robustness check. Comparing the oil-producing regions to the outcome of the regression tree analysis shows that all regions endowed with oil are combined in Group 3, except for West Kazakhstan[25]. Those regions also exhibit the highest group-average growth rate. This result is supported by Brunnschweiler (2008) and Ahrend (2002), who state that natural resource endowments have a positive effect on countries' real GDP growth and Russian region's economic performance, respectively.

However, this leaves us with the question of whether natural resource endowments are causally forming convergence clubs or whether our results are just a correlation. As natural resource endowment is taken as exogenous we would argue that the relationship can indeed be considered as causal. At the same time we, of course, cannot rule out that unobserved characteristics of resource-rich regions are crucial for the results.[26] To answer the question properly the interaction between natural resources and the variables important in determining economic growth would have to be analyzed in detail, which clearly leaves some space for further research. For the beginning it might be helpful to have a look at the relationship between natural resources and human capital as the number of infant schools per 1,000 persons turns out to be the most important split variable, with resource-rich

[24] Due to the small sample size it is not possible to formally check for parameter differences across the nodes like in Johnson and Takeyama (2001) .

[25] Among the oil regions, West Kazakhstan possesses, on average, the lowest share of oil, gas and coal production in total industrial production. This might be one of the reasons why West Kazakhstan is not included in Group 3.

[26] Variables like the degree of urbanization which are identified as being important in case of the Russian regions (Ahrend, 2002) are also considered in this study, but turn out to have no effect.

regions (and North Kazakhstan) having the most infant schools per 1,000 persons. This is especially surprising as most of the literature suggests a negative empirical relationship between natural resource abundance and the level of human capital.[27] In fact, looking at other human capital indicators for the Kazakh regions seems to support this. Apart from the infant schools, the resource-rich regions do, in general, not have the highest number of secondary schools, colleges and universities per 1,000 persons. However, the role of preschool education in determining high economic growth can reasonably be explained by the results of Heckman (2000). He points out that investment in preschool education gives the highest return to human capital investments.

[27] See for instance Gylfason (2001) and Gylfason and Zoega (2006).

5 Conclusion

The results of this case study confirm that for the Kazakh regions natural resources, at least partly, determine initial conditions at the beginning of the growth path and are thus important in forming convergence clubs. We further identify resource endowment to be more important than resource dependence and argue that this is due to the fact that resource endowment can be considered to be exogenous. Looking at the initial conditions, we find in addition that the resource-rich regions have more infant schools per 1,000 persons than the resource-poor ones. At first glance, this seems to be contradictory to most of the literature. Accounting for the relative importance of preschool education, however, makes our results plausible.

Nevertheless, the following concerns should be addressed. One problem is the small sample size. A higher number of observations would allow us to perform additional econometric analysis. Another issue could be the relatively short time period of 11 years as the convergence process might take more time.

Even though we are totally aware of these shortcomings, we are convinced that this paper is a valuable contribution to the empirical literature on natural resource abundance and club convergence in general and to the literature on the convergence process of the Kazakh regions in particular. Finally, we would also like our paper to be seen as a motivation for other researchers to test the hypothesis brought up in this paper in case of other regions or countries.

References

Agrawal, Pradeep, Economic Growth and Poverty Reduction: Evidence from Kazakhstan, Asian Development Review, 2008, 24(2), pp. 90–115.

Ahrend, Rudiger, Speed of Reform, Initial Conditions, Political Orientation or What ? Explaining Russian Regions Economic Performance, Working Paper 2002-10, DELTA 2002.

Ahrend, Rudiger, How to Sustain Growth in a Resource Based Economy? The Main Concepts and Their Application to the Russian Case, Working Paper 478, OECD 2006.

Aldashev, Alisher, Converging Wages, Diverging GRP: Directed Technical Change and Endogenous Growth, No. 307, OEI Working Papers 2011.

Azariadis, Costas and Allan Drazen, Threshold Externalities in Economic Development, The Quarterly Journal of Economics, 1990, 105(2), pp. 501–526.

Barro, Robert J., Human Capital and Growth, American Economic Review, 2001, 91(2), pp. 12–17.

Barro, Robert J., and Xavier Sala-i-Martin, Convergence, Journal of Political Economy, 1992, 100(2), pp. 223–251.

Barro, Robert J., and Xavier Sala-i-Martin, Economic Growth, 2nd Edition. The MIT Press, 2004.

Baumol, William J., Productivity Growth, Convergence, and Welfare: What the Long-run Data Show, American Economic Review, 1986, 76(5), pp. 1072–1085.

Bond, Stephen R. and Adeel Malik, Natural resources, export structure, and investment, Oxford Economic Papers, 2009, 61(4), pp. 675–702.

Breiman, Leo, Jerome Friedman, R. Olshen, and Charles J. Stone, Classification and Regression Trees, Monterey, CA: Wadsworth and Brooks, 1984.

Brunnschweiler, Christa N., Cursing the blessings? Natural Resource Abundance, Institutions, and Economic Growth, World Development, 2008, 36(3), pp. 399–419.

Bulte, Erwin H., Richard Damania, and Robert T. Deacon, Resource Intensity, Institutions, and Development, World Development, 2005, 33(7), pp. 1029–1044.

Canova, Fabio and Albert Marcet, The poor stay poor: Non-convergence across countries and regions, Working Paper 137, Department of Economics and Business, Universitat Pompeu Fabra 1999.

Carlino, Gerald A. and Leonard O. Mills, Are U.S. regional incomes converging? A time series analysis, Journal of Monetary Economics, 1993, 32(2), pp. 335–346.

Chambers, Dustin and Jang-Ting Guo, Natural Resources and Economic Growth: Some Theory and Evidence, Annals of Economics and Finance, 2009, 10(2), pp. 367–389.

Desdoigts, Alain, Patterns of Economic Development and the Formation of Clubs, Journal of Economic Growth, 1999, 4 (3), pp. 305–330.

Durlauf, Steven N. and Paul A. Johnson, Multiple Regimes and Cross-Country Growth Behaviour, Journal of Applied Econometrics, 1995, 10 (4), pp. 365–384.

Felipe, Jesus and Utsav Kumar, The Impact of Geography and Natural Resource Abundance on Growth in Central Asia, Working Paper, The Levy Economics Institute 2010.

Fischer, Manfred M. and Claudia Stirböck, Pan-European regional income growth and club-convergence, Annals of Regional Science, 2006, 40 (4), pp. 693–721.

Frey, Miriam and Carmen Wieslhuber, Do Kazakh Regions Converge?, No. 52, OEI Kurzanalysen und Informationen 2011.

Galor, Oded, Convergence? Inferences from Theoretical Models, Economic Journal, 1996, 106 (437), pp. 1056–1069.

Grier, Kevin B. and Gordon Tullock, An empirical analysis of cross-national economic growth, 1951–1980, Journal of Monetary Economics, 1989, 24 (2), pp. 259–276.

Gylfason, Thorvaldur, Natural resources, education, and economic development, European Economic Review, 2001, 45 (4–6), pp. 847–859.

Gylfason, Thorvaldur and Gyl Zoega, Natural Resources and Economic Growth: The Role of Investment, The World Economy, 2006, 29 (8), pp. 1091–1115.

Harris, Richard, Models of Regional Growth: Past, Present and Future, Journal of Economic Surveys, 2011, 25 (5), pp. 913–951.

Heckman, James J., Policies to foster human capital, Research in Economics, 2000, 54, pp. 3–56.

IMF, Macroeconomic Policy Frameworks for Resource-rich Developing Countries, Background Paper 2, International Monetary Fund 2012.

Isham, Jonathan, Michael Woolcock, Lant Pritchett, and Gwen Busby, The Varieties of Resource Experience: Natural Resource Export Structures and the Political Economy of Economic Growth, World Bank Economic Review, 2005, 19 (2), pp. 141–174.

Johnson, Paul A. and Lisa N. Takeyama, Initial conditions and economic growth in the U.S. states, European Economic Review, 2001, 45 (4–6), pp. 919–927.

Kutan, Ali M. and Michael L. Wyzan, Explaining the real exchange rate in Kazakhstan, 1996–2003: Is Kazakhstan vulnerable to the Dutch disease?, Economic Systems, 2005, 29 (2), pp. 242–255.

Lee, Kevin, M. Hashem Pesaran, and Ron Smith, Growth and Convergence in Multi-country Empirical Stochastic Solow Model, Journal of Applied Econometrics, 1997, 12 (4), pp. 357–392.

Meliciani, Valentina and Franco Peracchi, Convergence in per-capita GDP across European regions: A reappraisal, Empirical Economics, 2006, 31 (3), pp. 549–568.

Papyrakis, Elissaios and Reyer Gerlagh, The resource curse hypothesis and its transmission channels, Journal of Comparative Economics, 2004, 32 (1), pp. 181–193.

Papyrakis, Elissaios and Reyer Gerlagh, Resource abundance and economic growth in the United States, European Economic Review, 2007, 51 (4), pp. 1011–1039.

Roudoi, Andrei, Julia Zislin, and Bruce Bolnick, Kazakhstan Regional Disparities: Economic Performance by Oblast, No. 52, Nathan Associates Inc. 2011.

Siano, Rita De and D’Uva Marcella, Club Convergence in European Regions, Applied Economics Letters, 2006, 13 (9), pp. 569–574.

Siano, Rita De and D’Uva Marcella, Is there a Club Convergence among Italian Regions?, Working Papers 1, D.E.S. (Department of Economic Studies), University of Naples “Parthenope”, Italy 2006.

The Agency of Statistics of the Republic of Kazakhstan, Regional Statistical Yearbook of Kazakhstan 1996–1999 2000.

UNDP, NHDR 2009 – From exclusion to equality: Realising the rights of persons with disabilities in Kazakhstan., Human Development Report Office (HDRO), United Nations Development Programme (UNDP), 2009.

Ursulenko, Kseniia, Regional Development in Kazakhstan, No. 47, OEI Kurzanalysen und Informationen 2010.

Appendix

Table 5: Description of potential split variables

Mnemonic	Description	Average
AGRIEMP	employees in the agricultural sector as part of total employees	0.180521235
BIRTH	live births per 1,000 persons	15.41428571
BORROW	registered borrowers at public libraries as part of population	0.313941038
CHINFSCHO	children per infant school	114.2259603
COLLEGE	colleges per 1,000 persons	0.015700845
COLSTUDPOP	college students per 1,000 persons	9.013127431
CONSEMP	employees in the construction sector as part of total employees	0.041620336
DEATH	deaths per 1,000 persons	9.335714286
DOCTOR	doctors per 10,000 persons	29.92142857
FEMALE	female employees as part of total employees	0.427846448
HOSBED	hospital beds per 10,000 persons	80.88571429
INDEMP	employees in the industrial sector as part of total employees	0.24259426
INFSCHO	infant schools per 1,000 persons	0.096377054
INFSCHOPOP	children in infant schools per 1,000 persons	11.04608641
LANGUAGE	ratio Kazakh to Russian as language of instruction	2.38722426
LIBRARY	libraries per 1,000 persons	0.235705125
LIVSPACE	average living space per person (average of 1997 and 1999)	16.05
MURDER	murders per 100,000 persons	18.14285714
PUPSECSCHO	pupils per secondary school	429.6611475
PUTESECSC	pupils per teacher in secondary schools	12.29551389
SAVE	saving deposit per 1,000 persons	1.474832427
SECSCHO	secondary schools per 1,000 persons (1998/99)	0.565878112
SECSCHOPOP	pupils in secondary schools per person (1998/99)	0.209504947
STAAID	part of population receiving state aid (1999)	0.050260814
STIBIRTH	still births per 1,000 live births	22.41428571
STUDCOL	students per college	577.1036681
STUDUNI	students per university	2495.734524
SUICIDE	suicides per 100,000 persons	29.27857143
TRANSEMP	employees in the transportation sector as part of total employees	0.1110155
UNEMP	official unemployment rate	4.157142857
UNISTUDPOP	university students per 1,000 persons	14.88938578
UNIVERSITY	universities per 1,000 persons	0.006056923
URBAN	urban population as part of total population	0.53457183
VEHIC	registered motor vehicles per 1,000 persons (1999)	0.941854386