

STOCKHOLM SCHOOL OF ECONOMICS
Department of Economics
Course 659 - Bachelor's Thesis
Spring 2014

Modernisation - A Saviour Running Late

Assessing the validity of modernisation theory by investigating the effects of diversity on growth

Alexander Fält

22510@student.hhs.se

Adrian Rydén

22515@student.hhs.se

Abstract:

Modernisation theory is based on the idea that socioeconomic development leads to significant societal changes in culture and behaviour, changing the focus from group identity to individual ability. This paper seeks to investigate the validity of modernisation theory by examining the effects of ethnic, linguistic, and religious diversity on economic growth. This is done at the state level in the United States and India, using these as examples of countries that have reached different levels of development. Using OLS on a dataset constructed with official economic and census population data, we investigate the relationship between diversity and growth in the two countries separately. We find statistically significant and negative effects of diversity on economic growth in both countries, thus refuting the modernisation theory's claim that socioeconomic progress will diminish the importance of group categorisation along ethnic, linguistic and religious lines. The main contributions of this paper are therefore to demonstrate the shortcomings of modernisation theory in its classical form as well as to indicate a direction for future research in the study of economic effects of diversity at a subnational level.

Key Words: Modernisation, Diversity, Fractionalisation, Polarisation, Economic growth

JEL Codes: O40, O57, J10, R11

Supervisor: Örjan Sjöberg

Acknowledgements

We would like to express our sincere gratitude to our supervisor Örjan Sjöberg, Professor at the Department of Economics at the Stockholm School of Economics, for his valuable input and support during the process of writing this thesis.

We also wish to thank Göran Bostedt, Associate Professor at the Department of Educational Sciences at Mittuniversitetet, for providing guidance in the social sciences, comprehensible even to students of economics.

"The fusing process goes on as in a blast-furnace; one generation, a single year even - transforms the English, the German, the Irish emigrant into an American. Uniform institutions, ideas, language, the influence of the majority, bring us soon to a similar complexion; the individuality of the immigrant, almost even his traits of race and religion, fuse down in the democratic alembic like chips of brass thrown into the melting pot." - Titus Munson Coan, 1875

Table of contents

1. Introduction.....	3
2. Current state of knowledge.....	6
2.1 Terminology.....	6
2.2 Modernisation theory	6
2.3 Groups and behaviour - back to basics.....	7
2.4 Diversity and growth	9
2.5 Summary of current state of knowledge.....	11
3. Method	13
3.1 Fractionalisation and Polarisation	13
3.2 Ethnicity	15
3.3 Model specification	15
3.4 Heteroskedasticity	17
3.5 Hypothesis.....	17
4. Data	18
4.1 Availability.....	18
4.2 Diversity	18
4.3 Dependent variable	20
4.4 Controls.....	21
5. Results	24
6. Discussion	29
7. Conclusion	32
8. References.....	34
Appendix.....	36

1. Introduction

In an ever more globalised world where borders are becoming fluent and people are more mobile than ever, the subject of diversity and its economic effects has become increasingly present in the public debate. Whether a country benefits economically from diversity in its population is a question that has not only been discussed by politicians of different stripes but has also been taken up by the academic community. Empirical research into the topic has found diversity of some kinds to have an adverse effect on the economic growth of a country (Alesina & La Ferrara, 2005).

That diversity would be an issue today, especially in the highly developed parts of the world, is not in line with the predictions of the modernisation theory, an idea that can trace its roots back to the Age of Enlightenment (Inglehart & Welzel, 2005, p. 16). The theory is based on the concept of human progress. Its central insight is that socioeconomic development brings major social, cultural and political changes. As societies develop, they become more open and the relevance of social background and other ascribed characteristics decline (Marks, 2009). Many advanced postindustrial societies in the 21st century should therefore have reached a level of development where they would not see any of the potentially adverse effects of diversity.

The United States, a benchmark of wealth and development, is undoubtedly one of the countries that has come furthest in terms of socioeconomic development. It is widely thought that the country's entrepreneurial spirit, often considered to be a key factor behind its success, originates in the amalgamation of diverse groups that have immigrated to the country and keep knocking on its door. The country's ability to be a melting pot of cultures would seem to be a recipe for economic success and could serve as a validation of the predictions of modernisation theory regarding diversity.

A country that is ostensibly similar to the United States in many aspects is India. Except from partly sharing the US history as a former British colony, India is also a large democracy with a federal system of government. Like the US, or even more so, India is also a very diverse society with innumerable ethnic groups, languages and religions co-existing within the same subcontinent and under the same national flag.

Despite these, on paper, similar preconditions, India has not reached the potential that a diverse and federal democracy like the United States could be hypothesised to have inherent. With a GDP per capita, adjusted for purchasing power, at 5,138 \$, compared to 51,749 \$ for the United States in 2012, it is

obvious that they are very different countries.¹ This difference can of course be attributed to various factors; India's much longer period of colonialism, the destructive caste system or its history of planned economy to name a few.

A financial measure like GDP per capita cannot be said to entirely capture a country's level of development, since there are a number of other factors that have to be accounted for as well. Qatar, for instance, is a country that tops the list for this measure and it is hardly a paragon of modernity. Other factors such as institutional quality and education levels also have to be taken into account. Nevertheless, GDP can serve as a reasonable proxy and give an indication of how modern and developed a country is.

The implications for India, when considering its measure of GDP per capita, would be that its level of development is far below that of the US. When considering the economic effects of diversity for these countries we would then expect that, if there were a negative economic effect, India would see a more adverse effect than the US.

As a highly developed country, the kind of effect diversity would have in the United States is more ambiguous. The level of socioeconomic development required for diversity of different kinds to cease to be an issue is unclear. Empirical studies are contradictory as some find an adverse effect in the OECD countries (Patsiurko et al., 2013) while others find that sufficiently good institutions (which could plausibly be seen as a proxy for development) can mitigate the negative effect (Collier, 2000; Easterly, 2001; Tan, 2010). We will proceed under the assumption that the US, as a highly developed and postindustrial society, should, according to the implications of modernisation theory, not see the potentially adverse effects of diversity.

In this thesis we will make case studies of the United States and India, looking at the effects of different measures of diversity on economic performance. We investigate the validity of modernisation theory by studying the effects of diversity on growth. We aim to perform our study in a slightly new and different context by looking at a subnational level, that of US and Indian states. This approach keeps the level of development fixed to a large degree for the states as subnational entities within a country are assumed to be on similar levels of modernity. We will primarily look at the effects within the respective countries. Our research questions are as follows:

¹Data for 2012, gathered from the World Bank.

Is there a significant effect of diversity on economic growth? If so, is the effect of diversity dependent on a country's level of development?

The paper is organized as follows: Section 2 covers the previous research on our topic. In section 3 we review the method used to address the research questions and we formulate a hypothesis. Section 4 describes the data used in the study while section 5 presents the empirical results. These are then discussed in section 6 and lastly, in section 7, we present some concluding remarks.

2. Current state of knowledge

In this section we will outline the findings of previous research on our topic. We will start off this section by expanding our initial and brief introduction of modernisation theory. We will then provide a theoretical foundation, based in psychology and often lacking in empirical studies, to try to give an explanation as to why people who do not look alike, speak the same language or believe in different religions would have a harder time living together productively and peacefully. We then turn to the empirical findings and go through the different approaches to diversity that scholars have taken and summarize their findings. We conclude with a brief summary.

2.1 Terminology

Before delving into the theoretical framework that will be used a few words on the terminology would be appropriate. The term “diversity” is used to denote generic forms of divisions within a society. These divisions can take different shapes and various groups can have different kinds of identifying characteristics. These will be denoted when we want to be more specific, i.e. religious, linguistic, ethnic or racial diversity.

2.2 Modernisation theory

Modernisation theory is based on the idea of human progress. It is a relatively novel theory, as the idea of human progress did not seem very realistic in the steady-state equilibrium of agrarian economies. The theory itself originated in the Enlightenment era when the first explicit links between economic development and cultural change were proposed. Among those who made use of the ideas that underpin modernisation theory were both Adam Smith and Karl Marx. While they espoused very different beliefs in what society should look like, i.e. capitalism and communism respectively, they both saw technological innovation and its socioeconomic consequences as the basis of human progress. Since its inception and especially during the 20th century, the theory has gone through different periods of revisal, refutal and revival (Inglehart & Welzel, 2005, pp. 15-18).

Modernisation theory predicts that socioeconomic development brings systematic changes in political, social and cultural life. These changes translate into value shifts as society undergoes different phases of development and matters like traditional group membership become less important. This process leads to people becoming more secular, tolerant and trusting (p. 46).

Modernisation theory is not a theory in the conventional sense; it is more of a paradigm in which the belief that social background is becoming less important in the face of socioeconomic development is the unifying feature, although different social processes are emphasized (Marks, 2009).

A recently proposed revision of modernisation theory found that, while socioeconomic development and the rise of postindustrial society is linked to increasingly rational, tolerant and trusting values, cultural change is path dependent (Inglehart & Baker, 2000). Religion and ethnic traditions have proven to be surprisingly resilient throughout the world and a society's cultural heritage continues to affect the behaviour and values of its people (Inglehart & Welzel, 2005, p. 46).

While this revision might be better suited to explain the developments in different parts of the world, it does not seem to have been generally accepted as an updated version of the theory. Rather, we will use the more classical proposition that socioeconomic development changes society and shifts focus from group membership to individual traits.

2.3 Groups and behaviour - back to basics

Since our thesis will step outside the boundaries of rationality commonly assumed in economics (at least as rationality is usually understood), we will need to secure a very basic theoretical foothold of intergroup behaviour disregarding economic rationality. To do this we turn to social psychology to give some explanations as to why racial, ethnic, linguistic or religious diversity might affect the economic decisions of individuals. For group membership of various kinds to have any effect on people's actions there needs to be discrimination based on said group membership. For this discrimination to occur there must furthermore exist some sort of prejudice upon which to base the discrimination. (Hogg & Vaughan, 2010, p. 194). When prejudice is in place, discrimination may well follow, but there are many instances where prejudice alone does not lead to discrimination, at least not to its overt form (pp. 194-195). The causes of this prejudice have been the subject of much thought and empirical research. While early theories, originating in the middle of the twentieth century, sought explanations for prejudice in the individual and her personality and political outlook, focus later shifted to theories claiming that the formation of prejudice itself occurs in groups. Out of these latter theories we will use social identity theory, which has seen previous use in studies on diversity (Alesina & La Ferrara, 2005), but the reader should be aware that there are several other theories that could possibly explain the relevant intergroup dynamics (Hogg & Vaughan, 2010, p. 194).

Social identity theory argues that people have both social and individual identities, where the latter concern individual traits and interpersonal relationships and the former concern group traits and intergroup relationships. These two types of identities, united, then form a person's self. Social identities may stem from all types of group membership: national affinity, ethnicity, major subject of college studies and being in a whisky tasting club could all provide social identities (p. 215). The last example might seem trivial, but experiments have found that even temporary, randomly assigned group membership can cause significant ingroup favouritism, i.e. the mere existence of group divisions in an individual's mind can have important social repercussions (pp. 214-215). Whenever there exists a diversity of social groups, there exists a possibility that people will be categorised into these.

An explanation as to how this classification is performed is offered by self-identification theory in which individuals use prototypes, sets of attributes connected with different groups, to categorise themselves and others. Prototypes are based on the meta-contrast principle of "contrasts between contrasts", i.e. differences between different prototypes being greater than differences inside them. An example could be a person from a small town who moves to a large city to go to college. At home he primarily identified as an ambitious student or expert beer drinker, but in his new setting where all students are ambitious and no one drinks beer, he identifies with new prototypes such as those associated with being from his particular home region or having specialised in a particular subject in high school. When meeting exchange students he perhaps identifies as simply being from his home country and upon returning home he might then resume his original prototypes. Once a prototype is in place it is used to categorise both people within groups and outside of them. This causes individuals to push people around them into these groups and depersonalise them, "ingroupers" are thought to be more similar to oneself than what is probably the case and "outgroupers" are thought to be more different, emphasising differences between groups (Hogg & Vaughan, 2010, p. 216).

Connecting this back to social identity theory, we turn from the "how" to the "when" and "why" of identities and social categorisation. When categorisation is triggered depends on the accessibility and "fit" of the categories in question. Identities are activated when we use them frequently or when they "fit well with the cues in the present context". There are two reasons for having a social identity, self-enhancement and subjective uncertainty reduction. The former is based on differences in social status between groups causing individuals to try and identify with groups and compete with other groups to receive a positive social identity. The latter is based on a wish to reduce uncertainty, individuals want to know what to expect in different social situations and how to act when these situations arise (p. 216).

Because economic decisions are often made under uncertainty, the second reason is perhaps the most interesting in our case. If an individual is about to make an uncertain economic decision, she might turn to social categorisation and use relevant prototypes to reduce this uncertainty. If this means categorising others in such a way as to increase the estimated transaction costs for dealing with outgroup members, or reduce the estimated utility of decisions involving outgroup members, the aggregate economic effects could be significant if this categorisation is repeated by large shares of a population. It is worth noting the word subjective, meaning that objective uncertainty (for example the true probability of being tricked in a transaction) is not necessarily reduced by categorisation.

To establish a link with modernisation theory, we realise that, if we accept the premises of social identity theory, then modernisation must work in some way to reduce uncertainty (which is potentially harmful to the economy) in transactions if it is to have any effect on economic activities. Whether diversity has so far been shown to have an effect on economic performance will be the topic of the next section.

2.4 Diversity and growth

The subject of diversity and its effect on growth was brought to the attention of economists in the 1990's after a seminal paper on the "growth tragedy of Africa". Easterly and Levine (1997) showed that the growth rate of GDP per capita is negatively related to the degree of ethnic diversity within a country. They find that a large part of the growth differential between Africa and East Asia can be attributed to Africa's fragmented societies. Their article was inspired by an earlier paper on corruption that had used a measure of ethnolinguistic fractionalisation as an instrumental variable for corruption. In this paper Mauro (1995) argued that corruption causes slower growth and investment.

As a result of these studies a measure of ethnic diversity became a common control variable in regressions studying growth differences between countries (Alesina et al., 2003). The variable used was a measure based on ethnic fractionalisation², derived from data gathered by Soviet ethnographers in the 60's. This data used the origins of languages to categorise populations into different ethno-linguistic groups. This Ethno-Linguistic Fractionalisation (ELF) measure was duly criticized for its inability to distinguish between ethnic and linguistic differences, something that caused problems in parts of the world. In South America, for example, problems arose because people are much more homogenous in terms of language than in terms of ethnicity (Gören, 2014).

² Fractionalisation and how to measure it is further described in Section 3.1 but, in brief, it is based on population shares.

The criticism of ELF led to the creation of new indices of fractionalisation that make clear distinctions between ethnic, religious and linguistic heterogeneity (Alesina et al. 2003). These were further refined by trying to account for cultural distance between groups (Fearon, 2003; Bossert et al. 2011).

The studies conducted with these new measures confirmed earlier findings on the negative effects of ethnic fractionalisation and added observations about linguistic and religious fractionalisation. Linguistic fractionalisation was found to have a similar negative effect as its ethnic equivalent, not very surprising as language is part of the criterion to define the concept of ethnicity and the two are often strongly correlated. This inverse relation to growth observed for the ethnic and linguistic indices was however not found for religious fractionalisation (Alesina et al., 2003).

Following Easterly and Levine's seminal paper (1997), the question of institutional factors and their relevance to the effects of diversity was also raised. Collier (2000) finds that in the context of limited political rights, ethnic diversity is highly damaging to growth. Institutions seem to interact with ethnic diversity. Poor institutions have an even more adverse effect on growth in the presence of ethnic diversity. Levine (2001) argues that sufficiently good institutions can mitigate the negative effect. Tan (2010) finds that institutions and their interactions with ethnic diversity are central to explaining cross-country economic divergence. These observations are however not entirely consistent with the finding that ethnically homogenous and advanced capitalist countries tend to have stronger rates of growth than their ethnically diverse equivalents (Patsiurko et al., 2013).

An alternative way of measuring diversity has also been investigated. Polarisation³ was brought forward as a potentially better way to capture the conflict dimensions of diversity, which in turn can have an impact on economic growth. Montalvo and Reynal-Querol (2005) find that the measures were highly and positively correlated in homogenous countries. The correlation however drops drastically in highly diverse countries. As opposed to the direct effects that ethnic and linguistic fractionalisation had been observed to have on growth, they found polarisation for both ethnic and religious indices to have substantial indirect and adverse effects. These indirect effects work through an array of transmission channels. Gören (2014) finds that these channels include investment, civil war, human capital, government consumption, political instability, market distortions, trade openness and fertility when studying ethnic polarisation.

³ See Section 3.1 for further explanation.

The issue of potential endogeneity was raised since early studies used a single time-point for data on diversity while looking at extended periods for economic data (Alesina et al., 2003; Alesina & La Ferrara, 2005). When addressing this issue, ethnic fractionalisation as well as polarisation were found to be endogenous and subject to change over time. Campos and Kuzeyev (2007) find that treating diversity as dynamic leads to significant and negative results for ethnic diversity while the linguistic and religious kinds have no effect on growth. The effect of exogenous diversity on growth is found to be weak, a finding that is later also replicated by Campos et al. (2011).

Regarding the salience of diversity, different effects have been found for different kinds of diversity depending on the country being studied. In South Africa, racial fractionalisation appears to matter while other forms of diversity have little impact on economic development (Fedderke & Luiz, 2007). Alesina and La Ferrara (2005) find that the salience of diversity is also subject to change as the categorisation of groups is affected by economic and other policy choices.

Most studies to date have been performed on a cross-section of countries but Alesina & Zhuravskaya (2009) find that national levels of fractionalisation are often different from that of subnational levels. Dincer and Wang (2011) seemingly confirmed earlier findings at a subnational level as they find detrimental effects of both fractionalisation and polarisation on growth while treating diversity as dynamic, in a study on provincial level in China.

The potential benefits of diversity, such as its potential to enter the production function and give rise to complementary skill sets and increased productivity, have also been found in studies on both US cities and states. (Alesina & La Ferrara, 2005; Sparber, 2010). Ager and Brückner (2013) also, unlike most other studies, find a positive effect of fractionalisation, while conducting a study of the US in 1870-1920 and the effects of immigration. They find that when controlling for polarisation, which is found to be a key driver behind the negative effect of cultural diversity, increased fractionalisation is positive for growth.

Another issue that has been raised is the role of geography. Alesina et al. find it to be closely correlated with geographic variables like latitude. They show that ethnic fractionalisation is higher in countries closer to the equator. Tan (2010), however, dismisses the potential role for geography and Patsiurko et al. (2013) also find the suggestion that the size of nations play a part not to produce significant results.

2.5 Summary of current state of knowledge

Modernisation theory is based on the concept of human progress and it states that socioeconomic development bring major changes to society. Value shifts will occur as the development level rises and the focus will shift to individual traits and the importance of group belonging will decline.

To explain the negative effects of diversity often found in empirical research we turn to social identity theory, which traces the formation of prejudice and the discrimination that possibly follows to processes within groups. To reduce uncertainty and gain a positive social identity, people categorise themselves and others as members of groups defined by sets of attributes called prototypes. This leads to both ingroupers and outgroupers being ascribed attributes associated with their respective groups, to a larger extent than is actually reasonable. This leads to depersonalisation and an emphasis on intergroup differences.

Regarding the empirical findings, the economic effects of diversity have been studied from different angles, using different measures, methods and samples. The results of these studies have mostly pointed in the direction of diversity having a negative effect, although the magnitude of that effect and its significance has varied with the type of diversity in focus and the measure used. A number of issues have been brought up in the context of diversity and growth and the results have sometimes been at odds with one another.

3. Method

The purpose of this thesis is to explore the validity of modernisation theory by studying the effects of diversity on economic performance. To do so we must first establish definitions of the terminology that will be used, with terms such as fractionalisation, polarisation and ethnicity. We can then proceed to determine how to obtain measurements of the degree of fractionalisation and polarisation. To do this, as well as to determine what economic indicators should be used, we will draw upon previous literature. We will also discuss the control variables used to make a correct interpretation of our regressions possible and how this connects to the channels through which fractionalisation and polarisation are expected to affect economic performance. After this we will specify the regressions that we will and discuss how we will handle possible heteroskedasticity. We conclude with the formulation of a hypothesis.

3.1 Fractionalisation and Polarisation

Many of the articles discussed above make frequent use of the terms fractionalisation and polarisation. Here we will establish more clear definitions of these two terms along with ways of measuring them, which in the case of polarisation means both the optimal measurement and the measurement that we will have to settle for given our limited resources.

Fractionalisation

Fractionalisation is used in the literature as a synonym of diversity and fragmentation and so designates some sort of division in a society. The way of measuring fractionalisation in the literature is usually that the degree of fractionalisation in a country/state is the probability that a randomly selected person, belonging to some group (ethnic, linguistic, religious, etc.) belongs to a different group than a second randomly selected person in the same country/state (Easterly & Levine, 1997). The concrete way to measure this probability starts with using the Herfindahl index, which sums the squares of the population shares (Mauro, 1995):

$$H_i = \text{Herfindahl}_i = \sum_{j=1}^J n_{ij}^2$$

This calculates the Herfindahl index of fractionalisation along some group categorisation i using J groups and is equal to the probability of two randomly selected individuals belonging to the *same* group.

$\text{Max}(H_i) = 1$ for any number of groups as long as there is complete homogeneity (i.e. when $n_{ik} = 1$ for some $j=k$, $n_{ij} = 0$ for $j \neq k$), $\text{Min}(H_i)$ occurs when all groups are equal in size⁴, $n_{ij} = \frac{1}{J_i}$ resulting in

$$\text{Min}(H_i) = \sum_{j=1}^J n_{ij}^2 = J_i * \frac{1}{J_i^2} = \frac{1}{J_i}$$

This means that we have an index ranging from $\frac{1}{J_i}$ to 1. To get a measure of fractionalisation that increases when diversity increases we then define a fractionalisation index to be 1 minus the Herfindahl index. This will capture the probability of two randomly selected individuals belonging to two *different* groups and will range from 0 to $1 - \frac{1}{J_i}$

$$\text{FRACT}_i = 1 - H_i = 1 - \sum_{j=1}^J n_{ij}^2$$

Since the number of groups that are measured varies according to what is measured and in which country, it is important to remember what the index captures. The correct interpretation is that the index measures the probability that two randomly selected individuals belong to two different groups *that are being recorded*. Because all of our group categorisations include groups such as “others”, “two or more races” and “not reported”, the true fractionalisation will always be more or less underestimated.

Polarisation

As previously mentioned, polarisation is another measure of divisions within a society. Polarisation indices try to capture tensions and the risk of conflict between different ethnic, religious or lingual groups. Unlike fractionalisation this measure is maximised when there are two groups of equal size. Although polarisation is not the focus of this thesis, some authors argue that it is more important than fractionalisation when it comes to explaining the indirect effects of ethnic diversity on economic growth, using theories on rent-seeking. The ease of calculating it using the same data used for the fractionalisation measures allows a diversion into this as well. The index is built using groups' shares of populations as well as the distance between different groups. However, since such distances are very difficult to calculate, e.g. between ethnic groups, they are often assumed to be equal to allow polarisation indices to be calculated at all (Montalvo & Reynal-Querol, 2005). This is the imperfection described above. All of our polarisation measures use equal distances between groups. The resulting index ranges between 0 and 1:

⁴ This follows from the properties of the quadratic function $f(x)=x^2$. If we start with equally large groups, because $df/dx=2x$, the group that becomes larger will add more to the Herfindahl index than the group that becomes smaller subtracts.

$$Pol_i = 1 - \sum_{j=1}^J \left(\frac{0.5 - n_{ij}}{0.5} \right)^2 n_{ij}$$

3.2 Ethnicity

While some types of diversity can be readily defined, such as the religious or linguistic kinds, ethnicity is a more difficult matter. This is because the definition of a measure of ethnic diversity depends on the definition of ethnicity itself. There is also the added difficulty of how to separate so-called “racial” and ethnic measurements from each other. When determining how to proceed in constructing our ethnic variables, we will therefore draw upon previous literature and the ways that the census bureaus themselves categorise this. This will be discussed in further detail in the data section.

Endogeneity

The issues of endogeneity and the dynamics of diversity have been raised in previous studies. An example of the potential endogeneity that has been taken up by researchers concerns the self-reporting of ethnic identity in censuses being endogenous to government policies. If certain ethnic groups receive preferential treatment by the government then this might affect individual ethnic categorisation as people try to benefit from this preferential treatment. (Alesina and La Ferrara 2005). We furthermore acknowledge that diversity is potentially endogenous over time for other reasons but in our time periods this effect should only have a minor impact. The fact that the censuses are decadal also prohibits us from using more frequent data on total population, and we do not find that relying on estimates or smaller samples would add to the credibility of our results.

3.3 Model specification

To estimate the economic effects of diversity on growth we will perform regressions using ordinary least squares. We will use one type of diversity as an explanatory variable in each of these regressions and focus on these in our analysis.⁵ We will estimate the following linear model:

$$Growth_i = \beta_0 + \beta_1 Diversity_i + \gamma X_i + u_i$$

⁵ While including both fractionalisation and polarisation would be interesting and has been done in previous research (Ager & Brückner, 2013), multicollinearity issues among our observations prevented us from drawing any conclusions when trying this in our study.

The dependent variable is the average growth rate in state i during one growth period. $Diversity_i$ is one of the diversity indices, either in fractionalisation or polarisation form, and X_i represents the vector of control variables that we will include to reduce possible omitted variable bias.

Because we have fairly few observations for both countries it might be difficult to obtain statistical significance. For this reason, and because we are not really interested in the channels of diversity's effect on growth, we will attempt to minimise the number of controls used, to capture both direct and indirect effects of diversity and avoid the risks of overcontrolling. Furthermore, because we are looking at the within-country variation, some factors that are often controlled for in cross-country regressions will be differenced out, e.g. history and some types of geographical information (Ager & Bruckner, 2013). Choosing controls, we draw upon previous literature (e.g. Dincer & Wang, 2007; Patsiurko et al., 2013). All previous studies control for a convergence effect in per capita real economic activity, and so we include measures of the initial values for each period used (e.g. Easterly & Levine, 1997). This convergence is nonlinear, however, and we therefore include the log of initial income as well as the squared log of initial income for all growth periods (Patsiurko et al., 2013).

The only other control variable to be included in regressions for both countries is some measure of education level. This is because different population groups may have different educational attainment levels. In the US in 2003, for instance, 11,4 % of Hispanics had a bachelor's degree or higher, while 49,8 % of Asians had a bachelor's degree or higher (US Census Bureau, 2004). While we do not want diversity to absorb the effect of all education on growth, we also do not want to overcontrol. We believe that this must be considered a risk when including measures based on publicly funded educational institutions, because public education is just the type of public goods provision that previous cross-country studies have found people in diverse societies to be less willing to pay for. Therefore, we will use the percentage of the population that has completed the lowest level of non-publicly funded schooling as the educational control variable.

We will run the regressions for all types of diversity and for both countries. In the United States we will also have two different time periods. Once this has been done, we will run statistical inference to determine which, if any, of the types of diversity have a statistically significant impact on growth.

Much of the research conducted in this field has used an econometric technique called Seemingly Unrelated Regressions (SUR). This is a regression framework in which several equations are set up where the errors are allowed to be correlated across equations. This is often done when the effects of how

diversity works through specific channels are tested (Gören, 2014). However, we are not interested in specific channels, but rather in whether a society has learned how to overcome the difficulties of ethnic, lingual and religious diversity, whatever these difficulties may be and independent of whether the effects are direct or indirect.

3.4 Heteroskedasticity

When conducting our regressions we will, as always, have to be wary of potential heteroskedasticity. While robust standard errors are easily extracted in statistical software, there are potential hazards in using robust standard errors “by default” given our small number of observations, particularly for India where we only have 35 states to serve as potential observations. While the robust standard errors converge to the correct values in the limit when the number of observations goes to infinity, they can be severely biased when the number of observations is small, particularly if the errors are in fact homoskedastic (Söderbom, 2011; Wooldridge, p. 266). Instead of reporting robust standard errors in the results, we will test all regressions for heteroskedasticity and report relevant p-values for these tests. Should these tests confirm the presence of heteroskedasticity in any regression, those regressions can be found in the appendix with robust standard errors.

3.5 Hypothesis

Our hypothesis is derived from the implications of modernisation theory. As societies reach higher levels of development, there is a value shift that leads to people becoming more secular, tolerant and trusting (Inglehart & Welzel, 2005, pp. 46-47). In other words, the influence of social background and other ascribed characteristics decline (Marks, 2009). The implications for diversity should therefore be that it decreases in importance for economic growth as societies reach higher levels of development and that in a postindustrial society, the negative effect often found in cross-country studies, should not be observed. Our hypothesis, which will be tested for all possible regression configurations, is therefore as follows:

H0: Diversity has no effect on economic growth.

H1: Diversity has an effect on economic growth.

Thus, we expect to reject H0 in the Indian but not in the US regressions.

4. Data

4.1 Availability

While data is usually easily accessible for a host of variables at a country level, not only from the countries themselves but also from international organisations like the UN, the World Bank and the IMF, the same does not apply for data on state level. It has been surprisingly difficult to find data for our study and this has inhibited our ability to add control variables as well as having comparable data for longer periods of time, especially for India. Our lack of financial resources has also prohibited us from accessing some of the databases, particularly for India, that could have provided useful data. One consequence of this has been that we only look at one period for India while we use two periods for the US, which has to a greater degree met our data needs. The use of a single period for India had some positive effects as it increases our number of observations because of the addition of new states that were added in 2001 as well as the addition of Jammu & Kashmir, where the census was not conducted in 1991.⁶

4.2 Diversity

The primary sources for data on diversity are the censuses of the US and India. These are conducted once a decade and data from the United States has been gathered from the censuses of 1990 and 2000 while the census of 2001 has been used for Indian data. The diversity variables are thus of a single time point nature for each regression. We have used the census bureaus' categorisations when computing groups for all applicable measures of diversity. These censuses do not, however, gather data for all the different aspects of diversity that we are planning to look at, i.e. ethnicity, religion and language.

Starting with the US, there are no questions on religious affiliation, nor are there any questions on ethnicity, other than "Hispanic or Non-Hispanic". Regarding data on religious affiliation in the US, we have therefore made use of an alternative source. We have used data from a paper on segregation in the US, which also looks at the subnational level (Alesina & Zhuravskaya, 2009). Its data on religion in the US is based on a survey from the American Religious Identification Survey in 2001. This survey was conducted in the 48 contiguous states and the District of Columbia. Since the paper only gathered data for years around 2000 for its study, we only use their religious data for our regressions for the second period in the US.

With regard to ethnicity, we have followed the standard procedure in previous studies by using the racial data gathered in the US census to form our ethnic variables. This comes back to the salience of groups,

⁶ Civil unrest in the state prohibited the census of that year.

what characteristics are important differ among countries. In the US, like in South Africa (Fedderke & Luiz, 2007), the reasoning is that race is deemed a salient form of division and could be compared to the ethnic divisions in other countries. This has been computed in line with previous literature, drawing out Hispanic from the rest of the racial groups and giving it the status of a race in the calculation of the ethnic/racial index (Alesina & Zhuravskaya, 2009).

Another issue that arose for the US was that in the census of 2000 there was an addition of alternatives for race. This census gave individuals the option to choose to describe one's race with "Two or more" (US Census Bureau, 2000). This causes differences for the racial data between the periods and makes direct comparability very difficult. We will follow the procedure in the article on segregation (Alesina & Zhuravskaya, 2009) and choose to incorporate these new options into "Others".

Because of the addition of alternatives for race as well as our lack of data on religion for the first period, we are unable to construct a panel data set with the two US periods. We will therefore look at the two periods separately. For the US we will then first use linguistic and ethnic (racial) data from the census and in the second period we will make an addition of religious data taken from a dataset gathered for another paper.

The Indian census causes problems for other diversity variables. While the data for language (mother tongue and religion is readily available, there are no questions on ethnicity in the Indian census either. There are questions on caste but these only cover "Scheduled Tribes" and "Scheduled Castes", which have special protection in the Indian constitution and are often target for programs of affirmative action (Gang et al., 2008). We have constructed measures of diversity from these data as well but since they cause the remaining population to be classified as "others" while computing the measures, their use as a proxy for ethnicity seems limited. We will however once again follow Alesina and Zhuravskaya's (2009) lead and use it in that way while remembering the sort of data that underlies it.

While on the topic of "others", it should be noted that this category appears in both countries and in all kinds of diversity measures. The category "others", or in the case of US languages in the year 2000, "not reported", captures the remainder of, for example, linguistic groups. We have chosen to include this as a separate group in all measures of diversity, combining it with "two or more races" in the US 2000 racial fractionalisation measure. The nature of the Herfindahl index (adding squared shares) means that adding the "others" category mainly decreases concentration, i.e. makes the contribution to the index of any dominant group smaller. This leads to a fractionalisation measure between that given by excluding

“others” (using only the explicitly listed, separate groups as the total population) and the true one including all existing groups. In most cases the category is very small and does not significantly add to the Herfindahl index, but it does decrease the share of the dominant group. We do not believe that this will significantly influence our results.

As discussed in the method section, the ranges of different types of fractionalisation depend on the total number of groups used for that particular kind of fractionalisation. This firstly implies that “going from complete homogeneity” to “complete heterogeneity” means going from one group containing all individuals to all groups being equal in size. Secondly this means that to see the effect of going from complete homogeneity to complete heterogeneity in this sense, one must account for the value ranges that the fractionalisation measures are contained in. Since $0 \leq \text{Fract} \leq 1 - \frac{1}{J}$, if a fractionalisation variable has a coefficient of 1 and there are four groups measured this means that going from complete homogeneity to complete heterogeneity affects the dependent variable by 0.75.

4.3 Dependent variable

US

As previously mentioned, the dependent variable is the average growth of real GDP per capita. This has been averaged over different time periods, stretching as close as possible to ten years backward and forward from the measurements of diversity, although this has resulted in large variation in the time periods. Since the US Bureau of Economic Analysis (BEA) does not provide data on GDP per capita before 1987, we will use the average growth from 1987-2000 for our first period for the US and data from 1990-2010 for the second period. We wanted to maximize the years of growth around our data points on diversity without going over other census points as we also recognize the potential endogeneity in the diversity measures.

These periods includes a revision of GDP in 1997 as the BEA started using the NAICS system of calculation instead of the previous SIC system.⁷ We have chosen to include this change since the revisal of GDP presumably better captures the underlying growth in sectors such as fast-growing services and high-tech manufacturing sectors than was the case in the SIC system (BEA, 2007).

⁷ We tried both including the jump and bypassing it by taking the percentage increase from the revised 1997 figure when calculating growth for 1998. While the difference could be large for an individual state, it did not affect our results to any greater extent.

India

A similar issue in India is the frequent shift in base year and lack of data on GDP per capita. The Reserve Bank of India (RBI), which we deem to be a reliable source of data, provides data on NDP per capita for Indian states and we have therefore used that measure.⁸ To avoid changing base years, as they have very different and potentially huge effects across states, we used the data for NDP per capita between 1999/2000 to 2007/2008 as provided by the RBI. This keeps the data within the same base year and also allows for the use of data on the newer Indian states since their NDP per capita has also been computed backwards from their inception in 2001. It should be noted that data for three Union Territories is missing for our dependent variable, causing a drop in observations to 32. To be able to proceed we must assume that this does not cause a bias.

A secondary note of caution for the Indian data is the fact that RBI itself states that “Owing to difference in methodology of compilation, data for different States/Union Territories are not strictly comparable.” (RBI, 2013). As we have no more reliable source for data on Indian States, we have to make the assumption that the growth is sufficiently comparable and not biased in any systematic way.

4.4 Controls

Initial GDP is taken from the start of each growth period, i.e. 1987 and 1990 for the US states and the initial NDP for India’s states is taken for the year 1999/2000. The GDP data was taken from the BEA and the data on NDP was taken from the RBI.

The data on education for the US has been taken from the US censuses of 1990 and 2000 and is comprised of the percentage of the population in each state to have completed a bachelor’s degree or above. The data for Indian education has been taken from the census of 2001 and is comprised of the percentage of the population to have completed high school or above. This difference is attributed to the differing levels of educational attainment across the countries. It is also due to the fact that, to our understanding, it is only up to, but not including, secondary schooling that is provided free for all in India, as opposed to high school being the last level of free education in the US.

Dummy variable

⁸ Net Domestic Product (NDP) equals GDP minus depreciation for capital goods.

An initial investigation of the data revealed Alaska to be an outlier in the first period for the US in terms of growth.⁹ A probable reason for this is due to the starting year that we found ourselves using for our first period, 1987. This is close to the peak of oil production in Alaska and just before the decline started.¹⁰ Alaska is the only state with negative average growth between 1987-2000. Since this has a potentially large impact on the result, without being driven in any plausible way by diversity, we will include a dummy for Alaska in some regressions for our first period.

In our second period the decline in oil production had already started and presumably did not have an equally negative effect on growth, even though Alaska still has among the lowest growth rates for that period. With Alaska not being an outlier in this case, we will not use a dummy variable for the state in the second period.

Tables 1 and 2 provide additional description of the variables used in the regressions.

Table 1. Variable definitions for the US.

Variable	Definition	Obs.	Mean	Std.Dev.	Min	Max
Growth1987_2000	Average growth in real GDP per capita between 1987 and 2000 (dependent variable)	51	0.0251	0.00718	-0.0103	0.0353
Growth1990_2010	Average growth in real GDP per capita between 1990 and 2000 (dependent variable)	51	0.0167	0.00557	0.00159	0.0342
Log_of_Income87	Log of real GDP in 1987	51	10.24	0.271	9.803	11.37
Log_of_Income90	Log of real GDP in 1990	51	10.31	0.270	9.920	11.50
Log_of_Income_Squared87	Squared log of real GDP in 1987	51	105.0	5.695	96.10	129.2
Log_of_Income_Squared90	Squared log of real GDP in 1990	51	106.3	5.725	98.41	132.3
Education	Percentage of population to have achieved a bachelor's degree or above in 1990 and 2000	102	0.220	0.0489	0.123	0.391
Linguistic_Fract	Fractionalisation index for linguistic groups in 1990 and 2000	102	0.198	0.126	0.0496	0.566
Linguistic_Pol	Polarisation index for linguistic groups in 1990 and 2000	102	0.338	0.187	0.0962	0.8311
Ethnic_Racial_Fract	Fractionalisation index for racial groups in 1990 and 2000	102	0.333	0.163	0.0371	0.735
Ethnic_Racial_Pol	Polarisation index for linguistic groups in 1990 and 2000	102	0.544	0.229	0.0727	0.914
Religious_Fract	Fractionalisation index for religious	49	0.809	0.0584	0.623	0.874

⁹ See Appendix, Histogram 1, on the distribution of observations.

¹⁰ See Appendix, Graph 1, on the development of Alaskan oil production.

Table 2. Variable definitions for India.

Variable	Definition	Obs.	Mean	Std.Dev.	Min	Max
Average_Growth	Average growth in NDP per capita between 99-00 and 07-08 (dependent variable)	32	1.050	0.0182	1.011	1.087
Log_of_Income	Log of NDP in 99-00	32	5.198	1.701	2.035	7.683
Log_of_Income_Squared	Squared log of NDP 99-00	32	29.82	16.73	4.140	59.03
Education	Share of population that have educational attainment of high school or above in 2001	35	0.178	0.0704	0.0986	0.381
Ethnic_Scheduled_Fract	Fractionalisation index for ethnic groups in 2001	35	0.345	0.139	0.0355	0.607
Ethnic_Scheduled_Pol	Polarisation index for ethnic groups in 2001	35	0.617	0.214	0.0709	0.917
Linguistic_Fract	Fractionalisation index for linguistic groups in 2001	35	0.432	0.230	0.0637	0.905
Linguistic_Pol	Polarisation index for linguistic groups in 2001	35	0.534	0.178	0.123	0.818
Religious_Fract	Fractionalisation index for religious groups in 2001	35	0.322	0.170	0.0871	0.733
Religious_Pol	Polarisation index for religious groups in 2001	35	0.536	0.229	0.169	0.921

5. Results

USA 1990 census

We begin by running regressions for the United States using the growth period 1987-2000 with fractionalisation, polarisation and education data from the 1990 census. Log of GDP and squared log of GDP are taken from 1987. The results using fractionalisation, with non-robust standard errors, are shown in Table 3.

Table 3. Growth in real GDP per capita, average 1987-2000.

VARIABLES	(1)	(2)	(3)	(4)
Log_of_Income87	-0.125 (0.136)	-0.152 (0.133)	-0.217** (0.105)	-0.240** (0.101)
Log_of_Income_Squared87	0.00512 (0.00645)	0.00646 (0.00630)	0.0101** (0.00499)	0.0113** (0.00480)
Education	0.0580* (0.0324)	0.0447 (0.0312)	0.0115 (0.0259)	-0.00278 (0.0247)
Linguistic_Fract	-0.00985 (0.00877)		-0.00910 (0.00668)	
Ethnic_Racial_Fract		-0.00597 (0.00614)		-0.00783* (0.00462)
Alaska			-0.0336*** (0.00574)	-0.0344*** (0.00569)
Constant	0.760 (0.716)	0.900 (0.700)	1.187** (0.551)	1.304** (0.530)
Observations	51	51	51	51
R-squared	0.305	0.300	0.605	0.613
Adjusted R-squared	0.244	0.234	0.561	0.571

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

As mentioned above, Alaska is an outlier and has a large impact on our results. Without a dummy for Alaska the coefficients for racial and lingual fractionalisation are negative but not significant. Including a dummy for Alaska reduces standard errors for the fractionalisation coefficients and makes the ethnic/racial fractionalisation variable negative and significant at the 10% level. No such significance is found for lingual fractionalisation.

When testing for heteroskedasticity for all regressions we find that, without the Alaska dummy, the errors are heteroskedastic with extremely low p-values ($p < 0.0001$ for both types of fractionalisation). Adding the Alaska dummy results in the heteroskedasticity test statistic to lose significance ($p > 0.47$ for all

specifications including Alaska). We provide the results of the regressions with heteroskedasticity-robust standard errors in the appendix for the regressions without the Alaska dummy.

The regressions for polarisation for the same period in the United States are organised in the same way and are presented in Table 4. Just as with fractionalisation, we see that Alaska has a significant effect on the regression. Without including a dummy for Alaska, polarisation shows no statistical significance. Including the Alaska dummy does not produce any significant results either. As is the case above, the test for heteroskedasticity is highly significant for both types of polarisation without including the Alaska dummy ($p < 0.0001$ for both kinds of polarisation). After including the Alaska dummy, the tests for heteroskedasticity lose their significance ($p > 0.55$). As before, we include the regressions without the Alaska dummy with robust standard errors in the appendix, Table 10.

Table 4. Growth in real GDP per capita, average 1987-2000

VARIABLES	(5)	(6)	(7)	(8)
Log_of_Income87	-0.136 (0.136)	-0.157 (0.133)	-0.230** (0.106)	-0.246** (0.101)
Log_of_Income_Squared87	0.00564 (0.00646)	0.00667 (0.00631)	0.0107** (0.00503)	0.0116** (0.00480)
Education	0.0555* (0.0327)	0.0438 (0.0314)	0.00837 (0.0264)	-0.00520 (0.0249)
Linguistic_Pol	-0.00481 (0.00576)		-0.00392 (0.00441)	
Ethnic_Racial_Pol		-0.00318 (0.00390)		-0.00492 (0.00293)
Alaska			-0.0336*** (0.00581)	-0.0347*** (0.00571)
Constant	0.820 (0.718)	0.925 (0.701)	1.256** (0.555)	1.335** (0.530)
Observations	51	51	51	51
R-squared	0.296	0.296	0.596	0.613
Adjusted R-squared	0.235	0.235	0.551	0.570

Standard errors in parentheses

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

USA 2000 census

We now look at the US census data from 2000 using the growth period 1990-2010 and GDP from 1990 for log of GDP and log of GDP squared. This time we also have Alesina's religious fractionalisation and polarisation data. As before we start with the fractionalisation regressions with non-robust standard errors. The results are displayed in Table 5. We see that ethnic/racial fractionalisation is the only fractionalisation

variable significant at the 10% level. The heteroskedasticity tests cannot reject the null hypothesis of homoskedasticity at any reasonable significance level (all p-values > 0.54).

Looking instead at the polarisation for the same time period in Table 6, we see no significance for any type of polarisation. The heteroskedasticity tests for the regressions all have p-values > 0.38.

Table 5. Growth in real GDP per capita, average 1990-2010.

VARIABLES	(9)	(10)	(11)	(12)
Log_of_Income90	-0.234** (0.103)	-0.233** (0.0984)	-0.220** (0.0986)	-0.256** (0.104)
Log_of_Income_Squared90	0.0105** (0.00487)	0.0106** (0.00462)	0.0100** (0.00463)	0.0115** (0.00489)
Education	0.0580*** (0.0203)	0.0488** (0.0194)	0.0357* (0.0206)	0.0520** (0.0207)
Linguistic_Fract	-0.00625 (0.00600)			-0.00860 (0.00623)
Ethnic_Racial_Fract		-0.00913* (0.00455)		
Religious_Fract			-0.00215 (0.0122)	
Constant	1.297** (0.549)	1.287** (0.522)	1.205** (0.522)	1.400** (0.551)
Observations	51	51	49	51
R-squared	0.285	0.327	0.153	0.310
Adjusted R-square	0.222	0.268	0.076	0.233

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Table 6. Growth in real GDP per capita, average 1990-2010.

VARIABLES	(13)	(14)	(15)	(16)
Log_of_Income90	-0.240** (0.104)	-0.253** (0.0994)	-0.215** (0.0975)	-0.259** (0.104)
Log_of_Income_Squared90	0.0108** (0.00491)	0.0114** (0.00468)	0.00987** (0.00458)	0.0116** (0.00491)
Education	0.0574*** (0.0206)	0.0496** (0.0198)	0.0344 (0.0205)	0.0519** (0.0208)
Linguistic_Pol	-0.00316 (0.00429)			-0.00592 (0.00471)
Ethnic_Racial_Pol		-0.00508 (0.00342)		
Religious_Pol			0.00490 (0.00920)	
Constant	1.330** (0.554)	1.395** (0.527)	1.177** (0.519)	1.414** (0.553)
Observations	51	51	49	51
R-squared	0.276	0.301	0.158	0.305
Adjusted R-squared	0.213	0.240	0.081	0.228

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

India 2001 census

We now turn to India, using census data from the 2001 Indian census and average growth over the period 1999/2000-2007/2008. The results for the fractionalisation regressions are shown in Table 7. We see that only linguistic fractionalisation is statistically significant, at the 10% level. As with the United States, we perform heteroskedasticity tests for all regressions. These all have p-values of over 0.54, meaning that we cannot reject the null hypothesis of homoskedasticity.

Lastly we regress average growth on polarisation, the outcome is shown in Table 8. We now obtain significance at the 10% level only for religious polarisation. We also conduct the heteroskedasticity test, with p-values >0.50 for all three regressions.

Table 7. Growth in real NDP per capita, average 1999/2000-2007/2008.

VARIABLES	(17)	(18)	(19)
Log_of_Income99	0.00180 (0.0141)	0.00291 (0.0146)	0.00417 (0.0151)
Log_of_Income_Squared99	-0.000412 (0.00141)	-0.000334 (0.00147)	-0.000412 (0.00152)
Education	0.0428 (0.0457)	0.0651 (0.0467)	0.0757 (0.0538)
Linguistic_Fract	-0.0320* (0.0180)		
Religious_Fract		-0.0266 (0.0216)	
Ethnic_Scheduled_Fract			0.0214 (0.0300)
Constant	0.0592 (0.0372)	0.0425 (0.0357)	0.0198 (0.0327)
Observations	32	32	32
R-squared	0.163	0.115	0.083
Adjusted R-Squared	0.039	-0.016	-0.053

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Table 8. Growth in real NDP per capita, average 1999/2000-2007/2008.

VARIABLES	(1)	(2)	(3)
Log_of_Income99	0.00817 (0.0146)	0.00213 (0.0142)	0.00522 (0.0148)
Log_of_Income_Squared99	-0.000820 (0.00149)	-0.000234 (0.00143)	-0.000515 (0.00150)
Education	0.0566 (0.0473)	0.0730 (0.0460)	0.0703 (0.0512)
Linguistic_Pol	-0.0104 (0.0198)		
Religious_Pol		-0.0266* (0.0154)	
Ethnic_Scheduled_Pol			0.0125 (0.0188)
Constant	0.0276 (0.0338)	0.0481 (0.0344)	0.0181 (0.0332)
Observations	32	32	32
R-squared	0.076	0.159	0.081
Adjusted R-squared	-0.061	0.034	-0.055

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

6. Discussion

When interpreting our results we should be careful not to draw definitive and far-reaching conclusions. We are dealing with few observations and relatively short time spans for the average growth variables in both countries and although some measures of diversity show statistical significance, we quote: “As is well known, in many cases the results of cross-country regressions are sensitive to econometric specification.” (Alesina et al., 2003). While we do not deal with a cross-country study, we have experienced the same difficulty.¹¹ These factors give us reason to be cautious in our analysis. We do however believe that the results show tendencies that point in a certain direction. This, in combination with the instances that we reject the null hypothesis for different regression specifications, will help us answer our research questions regarding the effect of modernisation in the next section.

United States

Table 1 and column 4 imply that, during the first period for the US, going from complete ethnic/racial homogeneity to complete heterogeneity (equally large groups) would have an effect of -0.65% on the average yearly growth rate. The same coefficient for the second period is slightly more negative and, for both periods the coefficients are statistically significant at the 10% level. A negative effect is thus consistently found for measures of ethnic diversity in the US, a postindustrial and highly developed society that should, by the predictions of modernisation theory, have gone beyond the point where matters like racial group belonging should matter. For reasons previously mentioned, we do not consider these results very reliable, but enough evidence points in a negative direction, even if greater statistical significance would have strengthened our findings.

While there is no statistically significant effect of linguistic fractionalisation in the United States, the results still point in a negative direction. We must also recognise that especially the linguistic fractionalisation is prone to influence from recent migration, which possibly draws more people to faster-growing states. This could possibly induce an omitted variable bias, without which linguistic fractionalisation could also be significant and negative, although this is more speculation on our part.

¹¹ We have experimented with different specifications and in most cases our findings have been strengthened, e.g. when excluding squared log of income and including geographic variables. Given the current discussion about inequality, we also experimented with including a measure of inequality in the form of the Gini coefficient (to see if diversity was merely a proxy for inequality). This served only to increase the negative effect of diversity where such an effect was found but the data was only available for the US. We also found it hard to justify controlling for Gini, since we did not see any clear connection to modernisation theory.

India

The regression results for India make for a particularly interesting interpretation. Here we obtain statistically significant negative effects on growth for linguistic fractionalisation and religious polarisation and can reject H_0 at the 10 % significance level for these measures. These results could stem from the distribution of languages actually working in a different pattern compared to the distribution of religious affiliation to influence economic growth. When it comes to language, fractionalisation appears to be the salient measure of diversity because of its effects, direct or indirect, on growth. At the same time, religious fractionalisation might be less well suited than religious polarisation when it comes to picking up conflict along religious lines. The low correlation between linguistic fractionalisation and polarisation could be interpreted as lending support to the first conclusion. When it comes to religion, the correlation between fractionalisation and polarisation is much higher so it is hard to separate a polarisation effect from a fractionalisation effect.¹²

Meanwhile, ethnic polarisation and fractionalisation have positive coefficients. Though not significant, should they be found to have a positive effect over longer time periods this could be because of affirmative action efforts directed at the Scheduled Castes and Tribes, causing states in which this type of fractionalisation is high to receive growth benefits from these efforts.

When it comes to our research questions, the rejection of our null hypothesis for several kinds of diversity in both countries gives us an answer to the first one, whether diversity has a statistically significant effect on economic growth. On the second research question, whether the effect of diversity would be less important in the United States, our findings are not as clear. While the coefficients for the statistically significant kinds of diversity are larger in India, cross-country comparisons are not that simple to make. First of all, the coefficients contribute to the growth of different measures of real economic activity per capita, GDP and NDP respectively. Secondly, and more importantly, the coefficients should be viewed relative to the average growth of each country and the growth rates of India and the US are on very different levels¹³. Comparing the effect of maximum fractionalisation or maximum polarisation in the statistically significant cases divided by average growth could be a start for a cross-country comparison. Doing this for ethnic/racial fractionalisation we find fractionalisation/(average growth) effects of approximately 0.25 and 0.47 for the United States' first (excluding Alaska's growth) and second periods, respectively. In India these quotients are 0.60 and 0.53 for linguistic fractionalisation and religious polarisation, respectively. Comparing these numbers could have given us a hint as to the importance of

¹² See Appendix for tables on correlation

¹³ See mean of growth rates in Tables 1 and 2 in the previous section.

the relative effects of fractionalisation. While India exhibits somewhat larger quotients, there are no order-of-magnitude differences, which makes it hard to draw any reliable conclusions. Adding to the difficulty of making any cross-country comparison is the fact that India and the US are also extremely different in many other ways that our variables do not capture, the caste system in India being an obvious example.

Our findings are mostly in line with the previous research on diversity although our results are generally weaker, possibly due to our smaller number of observations and the fact that we do not use the SUR method and transmission channels. Similar to others, we also find consistently negative effects of diversity for both types of measures used. What differs from many other studies is that we find a measure of religious diversity, namely religious polarisation, to actually have a statistically significant and negative effect on economic growth. This could indicate that, while religious diversity is mostly found not to have an economically detrimental effect in cross-country studies, the situation could be different at the subnational level. Religion seems to be a salient form of division in India, something that could also hold for other countries at the subnational level or a subsample of countries if running cross-country regressions.

A potential weakness in our results is that they rely on the assumption that all states within the respective countries are on the same level of development, or modernity. Whether this is a credible assumption could be debated. There are states in both the US and India that differ greatly from the rest and the differences between states in India probably exceed those in the US. The capitals in both countries, where there are presumably greater concentrations of well-educated and well-paid people, are an example of this. Another example, which was partly treated as a dummy variable due to its outlier status, is Alaska. The fact is, however, that we deal with states that share a federal umbrella and by extension some measure of redistribution policies that, to some extent, can mitigate differences in growth and development. Coupled with the inherent difficulty in separating the different levels of government from each other and their contribution to the level of development for each state we believe that the assumption is sufficiently weak not to endanger the credibility of our findings.

7. Conclusion

Modernisation theory, the validity of which we have sought to investigate, predicts that socioeconomic development brings major social, cultural and political changes. Our findings show that the theory in its classical form does not correctly predict the effect of diversity on growth as we find negative effects in the United States, a highly developed and postindustrial society. Although our findings do not contradict modernisation theory's implications when it comes to the effects of diversity in India, as it is on a lower level of development, it becomes clear that there is a need for revision. A proposed update of the theory was mentioned in section 2.1 and we believe that it could potentially have merit in explaining the results in the US as well. The proposition that the changes that modernisation theory predicts are path dependent on cultural heritage could explain the effects we find for diversity in the US when considering its history of segregation. The effects of such a system could, according to this proposed revision of the theory, be surprisingly resilient and thus still affect economic outcomes. Our results therefore reject the classical version of the theory and lend support to the proposed revised version.

Contribution

The major contribution of this thesis in the empirical study of the effects of diversity is running our regressions with different types of diversity and different measures (fractionalisation and polarisation) of these types at the subnational level. To our knowledge this has not been done previously. As we see in India, we have statistically significant effects of linguistic fractionalisation and religious polarisation but not from linguistic polarisation or religious fractionalisation. This highlights the need for future research.

Topics for future research

While the implications that this thesis has for modernisation theory have come some way in disproving the validity of classical modernisation theory, there are also other topics related to the theories we have used and the results that we have produced that merit further investigation. Starting with the remaining theory used, we consider connecting this field of study to psychology to investigate the causes for prejudice and discrimination as well as their economic effects a high priority. Analysing growth means analysing improvements in the very real living standards of very real people, the importance of which cannot be overstated. Eliminating the root cause, uncertainty resulting in prejudice, is probably outside both the power and expertise of economists today. Efforts should however be made to understand the connection between group formation and identification, and economic decisions to attempt to find ways of minimising uncertainty in economic transactions. In the absence of a more permanent change in people's attitudes, finding a way to overcome the adverse economic effects of diversity could be invaluable.

Because our results show a significant effect of diversity in a modern, Western country, we believe that this type of study should be replicated in other developed countries. While some cross-country studies indicate that sufficiently high institutional quality can alleviate the negative effects of diversity, our results for the United States are particularly worrisome. While these could have been driven by US segregation in the country's fairly recent past, the perceived tolerance of the entire Western world comes into question when the "melting pot" itself seems unable to handle diversity. However, more countries need to be investigated and growth over longer time periods needs to be covered if any conclusions of value are to be drawn.

The possible path dependency of modernisation coupled with our finding of different forms of diversity to be salient in two different countries means that we see a need for continuing to make these studies at the subnational level, particularly to explore further how the measures fractionalisation and polarisation might differ in their effects in different countries.

Policy implications

It is difficult to see any direct potential policy implications. Changing the composition of a state's population is neither desirable nor legitimate and is unlikely to lead to any positive results. Although much is being done already, our advice is to focus on modernising and trying to reach higher levels of development. While we did not find support for the classical version of modernisation theory, its revised version could have merit. Therefore we still advocate modernisation because it can potentially mitigate some of the negative effects of diversity, with the rest of the effects needing solutions on a case-by-case basis. Even if modernisation alone does not decrease the importance of these adverse effects, economic development would provide additional resources for education, which could be used to try and reduce uncertainty among future generations.

8. References

- Ager, P. & Brückner, M. (2013) Cultural Diversity and Economic Growth: Evidence from the US During the Age of Mass Migration, *European Economic Review*, 64, pp. 76-97.
- Alesina, A., Devleeschauwer, A., Easterly, W., Kurlat, S. & Wacziarg, R. (2003) Fractionalization, *Journal of Economic Growth*, 8, pp. 155-194.
- Alesina, A. & La Ferrara, E. (2005) Ethnic Diversity and Economic Performance, *Journal of Economic Literature*, 43(3), pp. 762-800.
- Alesina, A. & Zhuravskaya, E. (2011) Segregation and the Quality of Government in a Cross Section of Countries, *American Economic Review*, 101(5), pp. 1872-1911.
- Bossert, W., D'Ambrosio, C. & La Ferrara, E. (2011) A Generalized Index of Fractionalization, *Economica*, 78, pp. 723-750.
- Campos, N.F. & Kuzeyev, V.S. (2007) On the Dynamics of Ethnic Fractionalization, *American Journal of Political Science*, 51(3), pp. 620-639.
- Campos, N.F., Saleh, A. & Kuzeyev, V. (2011) Dynamic Ethnic Fractionalization and Growth, *The Journal of International Trade & Economic Development*, 20(2), pp. 129-152.
- Collier, P. (2000) Ethnicity, Politics and Economic Performance, *Economics and Politics*, 12(3), pp. 225-245.
- Dincer, O.C. & Wang, F. (2011) Ethnic Diversity and Economic Growth in China, *Journal of Economic Policy Reform*, 14(1), pp. 1-10.
- Easterly, W. & Levine, R. (1997) Africa's Growth Tragedy: Policies and Ethnic Divisions, *Quarterly Journal of Economics*, 112(4), pp. 1203-1250.
- Easterly, W. (2001) Can Institutions Resolve Ethnic Conflict, *Economic Development and Cultural Change*, 49(4), pp. 687-706.
- Fearon, J.D. (2003) Ethnic and Cultural Diversity by Country, *Journal of Economic Growth*, Vol. 8(2), pp. 195-222.
- Fedderke, J.W. & Luiz, J.M. (2007) Fractionalization and Long-Run Economic Growth: Webs and Direction of Association between the Economic and the Social - South Africa as a time series case study, *Applied Economics*, 39, pp. 1037-1052.
- Gang, I.N., Sen, K. & Yun, M-S. (2008) Poverty in Rural India: Caste and Tribe, *Review of Income Wealth*, 54(1), pp. 50-70.
- Gören, E. (2014) How Ethnic Diversity Affects Economic Growth, *World Development*, 59, pp. 275-297.
- Hogg, M. A., Vaughan, G.M. (2010) *Essentials of Social Psychology*, Harlow, Pearson Education Limited.

Inglehart, R. & Baker, W. E. (2000) Modernization, Cultural Change and the Persistence of Traditional Values, *American Sociological Review*, 65, pp. 19-51.

Inglehart, R. & Welzel, C. (2005) *Modernization, Cultural Change and Democracy*, 1st ed. Cambridge: Cambridge University Press.

Marks, G.N. (2009) Modernization Theory and Changes Over Time in the Reproduction of Socioeconomic Inequalities in Australia, *Social Forces*, 88(2), pp. 917-944.

Mauro, P. (1995) Corruption and growth, *Quarterly Journal of Economics*, 110, pp. 681-712.

Montalvo, J.G. & Reynal-Querol, M. (2005) Ethnic diversity and Economic Development, *Journal of Development Economics*, 76, pp. 293-323.

Patsiurko, N., Campbell, J.L. & Hall, J.A. (2013) Nation-State Size, Ethnic Diversity and Economic Performance in the Advanced Capitalist Countries, *New Political Economy*, 18(6) pp. 827-844.

Reserve Bank of India (RBI), Table 9: Per Capita Net State Domestic Product at Factor Cost - State-Wise (At Constant Prices). Available at: <http://www.rbi.org.in/scripts/PublicationsView.aspx?id=15130>, accessed 17/4-2014.

Sparber, C. (2010) Racial Diversity and Macroeconomic Productivity across US States and Cities, *Regional Studies*, 44(1), pp. 71-85.

Söderbom, M. (2011) Nonstandard Standard Error Issues: A Guide for the Practitioner, *Lecture 11 of Econometrics II*, Available at: <http://www.soderbom.net/metrix2/lec11.pdf>, accessed 29/4-2014.

Tan, C.M. (2010) No One True Path: Uncovering the Interplay between Geography, Institutions and Fractionalization in Economic Development, *Journal of Applied Econometrics*, 25(7), pp. 1100-1127.

US Census Bureau (2004), Educational Attainment in the United States: 2003, *Population Characteristics*. Available at: <http://www.census.gov/prod/2004pubs/p20-550.pdf>, accessed 4/5-2014.

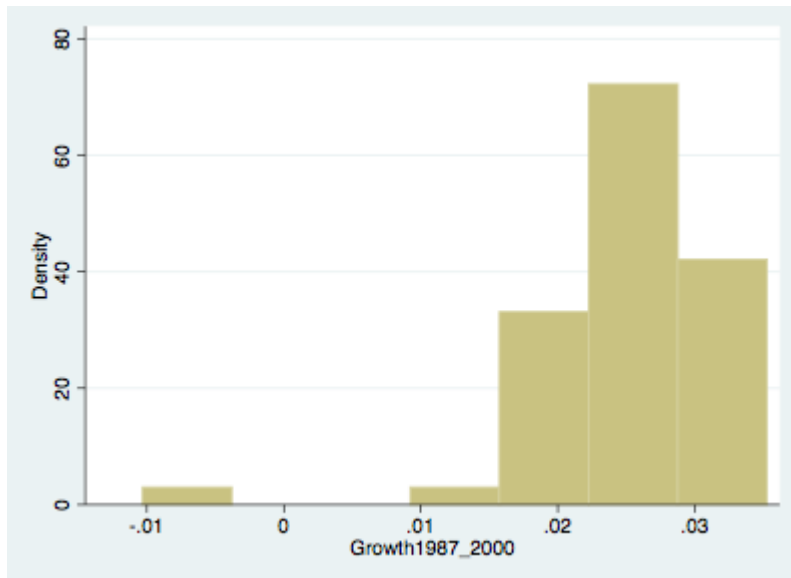
US Census Bureau (2000), Overview of Race and Hispanic Origin, *Census 2000 Brief*, Available at: <http://www.census.gov/prod/2001pubs/c2kbr01-1.pdf>, accessed at 6/5-2014.

Wooldridge, J. M., (2013) *Introductory Econometrics - A Modern Approach*. Fifth International Edition. Cincinnati: South-Western, Cengage Learning.

Yuskavage, W. (2007) Converting Historical Industry Time Series Data from SIC to NAICS, *Federal Committee on Statistical Methodology, 2007 Research Conference*. (BEA) Available at: http://www.bea.gov/papers/pdf/SIC_NAICS.pdf, accessed 5/5-2014.

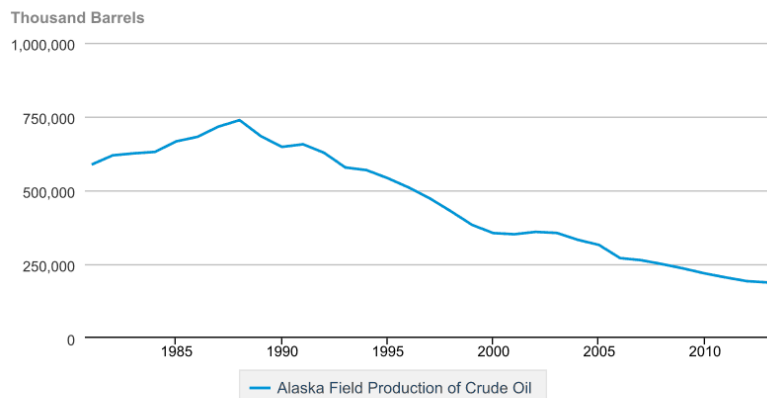
Appendix

Histogram 1. Average growth between 1987-2000 in US states, Alaska is identified as an outlier.



Graph 1. Alaska Field Production of Oil, *US Energy Information Administration*, available at: <http://www.eia.gov/dnav/pet/hist/LeafHandler.ashx?n=PET&s=MCRFPAK2&f=A>, accessed 10/5-14.

Alaska Field Production of Crude Oil



 Source: U.S. Energy Information Administration

Table 9. Growth in real GDP per capita, average 1987-2000

VARIABLES	(1) Growth	(2) Growth
Log_of_Income87	-0.125 (0.200)	-0.152 (0.203)
Log_of_Income_Squared87	0.00512 (0.00976)	0.00646 (0.00987)
Education	0.0580 (0.0446)	0.0447 (0.0435)
Linguistic_Fract	-0.00985 (0.00814)	
Ethnic_Racial_Fract		-0.00597 (0.00571)
Alaska		
Constant	0.760 (1.030)	0.900 (1.040)
Observations	51	51
R-squared	0.305	0.300

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Table 10. Growth in real GDP per capita, average 1987-2000

VARIABLES	(1) Growth	(2) Growth
Log_of_Income87	-0.136 (0.202)	-0.157 (0.204)
Log_of_Income_Squared87	0.00564 (0.00984)	0.00667 (0.00993)
Education	0.0555 (0.0448)	0.0438 (0.0442)
Linguistic_Pol	-0.00481 (0.00537)	
Ethnic_Racial_Pol		-0.00318 (0.00361)
Alaska		
Constant	0.820 (1.038)	0.925 (1.046)
Observations	51	51
R-squared	0.296	0.296

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Table 11. Country and fractionalisation type: Number of groups: possible value range of fractionalisation measure.

US racial 1990: 6: 0-0.83

US racial 2000: 7: 0-0.86

US linguistic 1990: 26: 0-0.96

US linguistic 2000: 324: 0-1

US religious 2000: 14: 0-0.93

India linguistic: 16: 0-0.94

India religious: 8: 0-0.88

India ethnic: 3: 0-0.67

Table 12. Correlation between diversity variables in India.

	LingFract	Relfract	EthFract	LingPol	RelPol	EthPol
Linguistic_Fract	1.0000					
Religious_Fract	0.4041	1.0000				
Ethnic_Scheduled_Fract	-0.0996	-0.0699	1.0000			
Linguistic_Pol	0.5617	0.1318	-0.0438	1.0000		
Religious_Pol	0.2999	0.9348	-0.1583	0.1865	1.0000	
Ethnic_Scheduled_Pol	-0.0872	0.0140	0.9633	-0.1081	-0.0874	1.0000

Table 13. Correlation between diversity variables in the first US period.

	LingFract	EthFract	LingPol	EthPol
Linguistic_Fract	1.0000			
Ethnic_Racial_Fract	0.6447	1.0000		
Linguistic_Pol	0.9415	0.6471	1.0000	
Ethnic_Racial_Pol	0.484	0.9564	0.4867	1.0000

Table 14. Correlation between diversity variables in the second US period.

	LingFract	RelFract	EthFract	LingPol	RelPol	EthPol
Linguistic_Fract	1.0000					
Religious_Fract	-0.0136	1.0000				
Ethnic_Racial_Fract	0.6368	-0.0418	1.0000			
Linguistic_Pol	0.8932	0.0132	0.6703	1.0000		
Religious_Pol	0.0227	-0.9278	-0.0120	-0.0146	1.0000	
Ethnic_Racial_Pol	0.5041	-0.1075	0.9689	0.5142	0.0298	1.0000