Is Diversity Bad for Economic Growth? Evidence from State-level Data in the US

Nazmun N. Ratna**
Lecturer in Economics
Lincoln University, Canterbury

R. Quentin Grafton
Crawford School of Economics and Government
The Australian National University (ANU), Canberra

Tom Kompas
Crawford School of Economics and Government
The Australian National University (ANU), Canberra

Abstract

The paper examines the macroeconomic effects of social diversity in the United States. Employing a cross-sectional dataset for 48 contiguous states in the US, we find empirical evidence for a negative impact of diversity on Gross State Product (GSP) per capita growth. The findings indicate that racial diversity has a negative economic impact in the absence of offsetting factors that would help to overcome barriers to communication across social groups. After controlling for low levels of English fluency, or the inability to communicate effectively, the estimated negative economic impact of racial diversity is even more pronounced. If low-level English fluency is positively correlated with the cost of communication across the population, as we would expect, the results indicate that racial diversity reduces GSP per capita growth when barriers to communication are higher. The results provide an economic justification for establishing 'weak ties' across diverse social groups in pluralistic societies.

***JEL* Classification: O40, C21, Z13**

**Corresponding Author:
Nazmun N. Ratna
Email: ratnan@lincoln.ac.nz

Presented at

Econometric Society Australasian Meeting ESAM08

Wellington, New Zealand

9-11 July 2008

Section I: Introduction

An emerging literature in growth economics explores the influence of social factors on the process of economic growth. Empirical evidences, principally from cross-country studies, suggest that social factors like, diversity, trust and social norms, influence aggregate economic performances. In this paper we identify racial diversity as a causal factor for differences in economic growth across 48 contiguous states in the US.

The results are interesting because they provide a more subtle or nuanced interpretation of the existing view that diversity has a positive impact on the US economy. Concentration of human capital in diverse cities and the consequent economic outcomes has been one of the core concept investigated as part of growth economics as well as economic geography literature. The principal argument is, a diverse society works as a breeding ground for new ideas or innovation due to complementarity in knowledge, skills or experiences, and the consequent knowledge spillovers. Recent studies extend previous findings with US datasets and identify diversity as a causal factor for economic indicators like hourly wage of native workers (Ottavianno and Peri 2005) or labour productivity in manufacturing industries (Sparber 2006). Our investigation makes a contribution by focusing on a comparatively unexplored subject area, namely, economic effects of barriers to communication or 'social divergence' (Grafton, Knowles, and Owen 2004), created through differences in language, ethnicity or religion. We hypothesize that social segregation in diverse societies retards the process of knowledge diffusion across the groups/networks and hence influence economic growth. Given the subtlety surrounding appropriate measures for social segregation, in a country like the US, a range of definitions has been employed in this investigation to capture different dimensions of social segregation. We find that racial diversity is the only statistically significant measure, contributing negatively towards Gross State Product (GSP) per capita growth.

In order to assess if the negative economic impact of the deeply rooted social segregation in the US has to do with barriers to communication, we explore a range of alternative hypotheses and check the validity of our estimate with an extensive set of control variables based on the existing literature. In this paper barriers to communication are defined as linguistic diversity coupled with low levels of English fluency. Our estimates show that when we control for low level of English fluency, or the inability to communicate effectively in an English-speaking society, the effect of racial diversity is stronger. If low level of English Fluency is positively correlated with the cost of communication across the population, the results support the view that the economic impact of racial diversity is negative when barriers of communication are higher.

The paper is structured as follows. In section II we explore the motivation and background of the literature on social diversity and its principal findings. Section III describes the data and variables used to examine the impacts of social diversity on state-level economic performance, section IV outlines the model and initial estimation results, section V provides a series of robustness checks on the findings and section VI an in-depth examination of the issue of endogeneity. Section VII provides a brief review of the policy implications while section VIII offers concluding remarks.

Section II: Motivation /Background

The principal themes regarding negative impact of diversity or ethno-linguistic fractionalization include: one, non-optimal economic decision-making due to lower level of interpersonal trust and two, conflict of preferences over redistributive policies. One of the pioneering studies on detrimental impact of ethnic diversity by Easterly and Levine (1997) argues that the public policy choices in ethnically fragmented societies are not economically optimal due to the conflict of preferences. Alesina et al.(2003) and also Alesina and Ferrara (2005) show that ethnic and linguistic fragmentation measures have a negative impact on per capita growth. The empirical evidence for sub-optimal policy choices for provision of public goods like education or public employment across different geographic boundaries in the US is given by Porteba (1997), Alesina et al. (1999), and also Goldin and Katz (1999).

More relevant for our analysis are the empirical evidences on the causal relationship between trust and macroeconomic indicators. Knack and Keefer (1997) argue that a higher level of trust ensures increased incentive for innovation, accumulation of human capital and better performance of government institutions. Zak and Knack (2001) hypothesize that a higher level of trust is associated with an egalitarian distribution of income and population homogeneity, which enhances investment and thereby economic growth. The principal argument is that the lower level of interpersonal trust[i] and civic engagement in a fragmented society influence the aggregate economic performance through reduced participation in collective action problem and an increasing transaction cost of communication across groups. Explaining the poor growth performance in Sub-Saharan Africa, Collier and Gunning(1999) identify lower level of trust or poor 'public social capital', as one of the major impediments.

Our investigation focuses on social divergence that argues that barriers to communication hinder the 'cross-fertilization' of ideas and knowledge due to lower social interactions across the groups and, hence, has a negative impact on productivity, and ultimately on factor accumulation (Grafton et. al. 2007). Lower level of interpersonal trust is documented in localities with higher racial heterogeneity and income inequality in the US (Alesina and Ferrara(2002) or the linguistically diverse neighbourhood in Australia ((Leigh 2006). Grafton et al.(2004) analyse the significance of social divergence in explaining cross-country differences in Total Factor Productivity (TFP), and per capita income. Based on an optimal growth model, Grafton et al. (2007) explore the theoretical basis for the principal hypothesis of social divergence. The authors also use a cross-country dataset of 110 countries and conclude that that the impact of linguistic diversity is negative for TFP growth.

The principal hypothesis of social divergence is not that social diversity is undesirable but that it makes communication more costly and, thus, inhibits the generation and dissemination productivity enhancing knowledge. Although not mutually exclusive, growth economics and economic geography or urban economics literature provide evidence that diversity contributes to innovation and knowledge by offering a range of abilities, experiences and cultures, which can potentially be complementary in nature. In terms of social diversity hypothesis, the gains from trade exceed the costs of trade, particularly, in large and economically vibrant cities. The creative aspects of cities like New York, London, Berlin or San Francisco, provide the rationale for positive economic impacts of heterogeneity that originate from the availability of a variety of skills and experiences, and concentration of human capital in a diverse city (Florida 2002).

Recent studies on diversity and economic performances, such as by Ottaviano and Peri (2005), provide evidence that linguistic diversity contributes positively to hourly wages, and employment density of US natives. The authors consider linguistic diversity as a proxy for cultural diversity, and argue that different skills originating from different cultures contribute to the productivity of native workers. Such a result, however, is not necessarily inconsistent with the social divergence perspective that barriers to communication across social groups have a negative impact on economic performance. In other words, provided that people can overcome existing social barriers to communication, such as, by both groups speaking well a mutually intelligible language, the gains in trade from communicating complementary knowledge more than offset the costs of communication. Two other studies that investigate to the effect of racial diversity on a set of macroeconomic indicators — labour productivity across US industries (Sparber 2006), GSP per worker at state level, and average wages of white workers at city level (Sparber 2006) — has been mixed. The impact of racial diversity proves to be positive for most industries in the US, more so for industries with high skill concentration, and productivity of white workers. To the extent that education and high levels of skill proxy lower costs of communication this result is also consistent with the social divergence hypothesis. State-level results of the impacts of racial diversity on economic outcomes are inconclusive ((Sparber 2006).

Section III: The Data and Variables

State-level, cross-sectional data for the 48 contiguous states in the US is used in our modelling. The data are from several sources. To construct the dependent variable, Gross State Product (GSP), data is obtained from the US Bureau of Economic Analysis for the period 1999-2000 (US Department of Commerce 2004); and is defined as the sum of value added originating in 63 industries in each state. GSP per capita, is measured as the ratio of real GSP[ii] and state population for 1999 and 2000.

To construct the key causal variables, the diversity measures, the data on language and race has been obtained from the US Census Bureau for 1990 and 2000. Two components of language data provided in the Census: *language spoken at home* and *ability to speak English,* has been used. Respondents were asked if the person 'speaks a language other than English at home' and the data are compiled for population 5 years and over, into five broad categories: English, Spanish, other Indo-European languages, Asian and Pacific Island Languages, and all other languages[iii]. The Census Bureau defines race as the self-identification data item, in which people choose the race(s) with which they most closely identify. It also specifies categories that basically represent both racial and national origin groups. We use the data on race from both the 1990 and 2000 census. For Census 2000, six categories were included for respondents reporting one race[iv]: i White; ii Black or African American; iii American Indian and Alaska native; iv Asian; v. Native Hawaiian and other Pacific Islander; and vi some other race. To measure religious diversity, the data from the Statistical Abstracts of the United States in 2000is used for Christians and Census 2000 for data on the Jewish population[v].

We consider three diversity measures to evaluate the social barriers to communication: racial diversity, linguistic diversity and religious diversity[vi]. All measures of diversity are derived as the fractionalisation index, a widely used measure in cross-country regressions:

$$\text{FRAC}_i = 1- \sum_j^n f_{ji}^2$$

Where f_{ji} is the share of group (racial, linguistic or birth region of foreign population) j (j=1, 2,…, n) in state i. It provides a *Herfindahl* concentration index and signifies the likelihood that two people chosen at random from a diverse universe, will belong to different groups[vii]. In addition to the fractionalisation index, for religious diversity the *Polarisation* index (Reynal-Queral 2002) has been used for religious diversity only. A polarization index addresses the issue of evenness from a different perspective (Maignan et. al. 2003). Based on the theoretical results of Esteban and Ray (1994 in Alesina and Ferrara (2005), Reynal-Querol index (RQ) is defined as:

$$\text{RQ} = 1- \sum_{j=1}^n (0.5 - p_j)^2 p_j / 0.25,$$

where p_j represents the share of group *j* in the population[viii]. The main reason for using a polarization index lies with the limitation of the data where only two religions are identified as a separate category. As richness is not incorporated in the data properly, the use of a polarization index enables us to evaluate evenness in a better manner compared to the fractionalisation index.

To evaluate the impact of other factors contributing to TFP, a set of control regressors are selected on the basis of existing literature on economic growth in the US. All of these are derived from the Census 2000 dataset from US Bureau of Census. Educational attainment *or education* is defined as the percent of the population with a Bachelor's degree or higher[ix], which serves as a proxy for human capital. A significant number of studies on the US economy have shown that differences in level of human capital is crucial for huge differences in per capita income or productivity across different geographical areas – states or counties (Beeson 1987; Ciccone and Hall 1996; Sedgley and Elmslie 2004; Higgins, Levy, and Young 2006; Hendricks 2004), and cities or metropolitan areas (Glaeser et al. 1992; Segal 1976; Rauch 1993; Shapiro 2006; Ciccone and Hall 1996; Simon and Nardinelli 2002; Rigby and Essletzbichler 2002). Although a number of competing hypotheses exist about the source of human capital externalities in cities or metropolitan areas, the empirical evidence for its linkage with economies of agglomeration or urbanization, has been strongly established in the case of the US economy. Hence, we include two agglomeration variables: *Urbanization*, defined as the percentage of population in metropolitan areas, and density of population or *PopDensity*. *Urbanization* is measured as the percentage of the state population living in metropolitan areas. Population density (*PopDensity)*, defined as the average number of inhabitants per square mile of land area, is from the Census 2000 dataset. Land area is the size, in square miles, of all areas designated as land in the Census Bureau's national geographic database. Population density enables knowledge spillovers (Marshall 1920, cited in Glaeser 1998; Fu 2007 and employment growth (Jacobs 1969 cited in Fu 2007; Glaeser et. al.(1992)). Inclusion of population density as a control regressor also enables us to verify the significance of 'weak ties' (Granovetter 1973) in information diffusion.

Given the findings in a study by Persson and Malmberg (1996), we also include the percentage of working age population, defined as *Age*, as a control variable. *Age* is the percentage of state population in the productive age group, from 20-64 years. The Persson and Malmberg (1996) study provides evidence that age structure influences growth of per capita income across the US states for the period 1920-1990.

Section IV: Impacts of diversities on US states

In this section, first, the preferred framework for evaluating the impact of social diversities has been explored. Then the robustness of the preferred framework has been checked in the following sections. Summary statistics for the key variables are presented in Table 1.

4.1 Estimation method: Level vs. growth

To identify social diversity as a causal factor for economic performance, in this paper a range of variants of equation (4.1) has been estimated, where the dependent variable is the natural logarithm of GSP per capita for the *i*-th state for a particular year *T*:

$$\ln GSP_{iT} = \mathrm{p} + \mathrm{a}\,Racial + \mathrm{b}\,Language + \mathrm{d}\ \mathrm{Re}\,ligious + \mathrm{g}\,Control_i + \mathrm{e}_i \qquad (4.1)$$

The explanatory variables in equation (4.1) consist of three measures of diversity - *Racial*, *Language* and *Religion*. Based on the existing literature on economic growth, C*ontrol,* is included as a vector of other explanatory variables such as education, urbanization, and population density.

In equation (4.2) GSP_{it} is the natural logarithm of real GSP per capita of the *i*-th state in some initial year *t*. The regressand in equation (4.2) thus represents growth in GSP per capita. The right hand side variables, however, are the same as the level framework in equation (4.1).

$$\frac{1}{T}\ln\left(\frac{GSP_{iT}}{GSP_{it}}\right) = \mathrm{p} + \mathrm{l}\ \ln GSP_{it} + \mathrm{a}\,Racial + \mathrm{b}\,Language + \mathrm{d}\ \mathrm{Re}\,ligious + \mathrm{g}\,Control_i + \mathrm{e}_i \quad (4.2)$$

Equation (4.2) is a framework has been widely applied in empirical growth economics. Surrounding the issues of conditional convergence, the principal argument for this framework is that the initial GSP captures the transitional dynamics when the economies under the study are not in their steady states (Mankiw, Romer, and Weil 1992).

To test for the validity of the Mankiw et al. (1992) approach and arrive at the growth specification we perform an omitted variable test as to whether initial GSP per capita has added explanatory power, or is statistically significant by estimating equation (4.3). Equation (4.3) is derived from rewriting equation (4.2) and the null hypothesis, H_0: $(1+\mathrm{l})=0$, reduces (3) to the level equation (4.1). The rejection of the null hypothesis i.e. H_0: $(1+\mathrm{l}) = 0$, provides evidence for the growth specification.

$$\ln GSP_{iT} = \mathrm{p} + (1+\mathrm{l})\ln GSP_{it} + \mathrm{a}\,Racial + \mathrm{b}\,Language + \mathrm{d}\ \mathrm{Re}\,ligious + \mathrm{g}\,Control_i + \mathrm{e}_i \quad (4.3)$$

The OLS estimates for equation 4.1 in column (1), Table 2 reveal that though linguistic and racial diversity indexes have the hypothesized sign, none of them is statistically significant. The coefficients for each of the control variables- *Education*, *Urbanization* and *Age*, however, have the hypothesized positive sign and are statistically significant at 5-percent and 1–percent level respectively. Given the correlation (0.533) between *Urbanization* and *PopDensity*, we re-estimate equation (4.1) with population density instead of urbanization. The results are presented in column (2). Although the coefficient for *PopDensity* is statistically significant at the 20-percent level (with a p-value of 0.116), the extent of influence is much smaller than that of urbanization. The statistical significance for *Education* increases, but the overall goodness of fitness reduces under this specification. *Urbanization*, therefore, has been used as the agglomeration variable for the rest of the regressions.

Column (3), Table 2 gives the result of equation (4.3), which includes initial GSP or *lnGSP99* as a regressor in growth specification. The null hypothesis i.e. $H_0: (1+\lambda) = 0$, is strongly rejected (with a t statistic of 38.342) at the 1-percent level and thus proves to favour the growth framework, as specified in equation (4.2). The inclusion of a lagged dependent variable in equation (4.1) is helpful on two grounds: first, to help determine the preferred framework, and then, to account for factors like resources, infrastructure and other historical factors, inclusion of which is constrained given the small data set. Inclusion of initial GSP, therefore, also addresses the potential for omitted variable bias (Woodridge 2003).

Column (3), Table 2 also provides an interesting result regarding the coefficient of racial diversity. It has the hypothesized negative sign, and is statistically significant at 5-percent level. This result supports our hypothesis that the racial division within a particular state creates barriers to communication and, hence, has a negative impact on economic growth due to restricted flow of knowledge or information. Linguistic diversity, however, has a positive sign and is statistically significant at the 5-percent level. Although not expected according to our hypothesis, the positive sign of linguistic diversity supports the findings of Ottaviano and Peri (2005). In a country like the US, this result does not seem to be surprising, as only 17.9% of the population (5 years and over) spoke a non-English language at home in 2000 (Shin and Bruno 2003). We elaborate issues concerning linguistic diversity in the later part of this section.

Our estimates confirm the expected positive sign of the coefficient for *education*, which is statistically significant at the 1-percent level. The reported diagnostics are *Normality*, i.e., χ^2 test for normality of errors and *Hetero*, i.e., an F-statistic to test for heteroskedasticity. In column 1, Table 2 the heteroskedasticity test statistic is not statistically significant. The *Normality* test statistic is, however, significant at 1-percent level and thus provides strong evidence for non-normality of errors. With the inclusion of initial GSP per capita, the null hypothesis cannot be rejected for both test statistics (column 3, Table 2) and overall goodness of fit also improves.

Column (4), Table 2 reports the OLS estimates of the growth framework specified in equation (4.2). Three more variables are added under this framework: *Poverty* and interaction terms for linguistic and racial diversity with net migration rate. *Poverty*, is defined as the percentage of families below the poverty line in the immediate past (last 12 months) in 2000, and is obtained from 'State and Metropolitan Area Data Book' (U.S. Department of Commerce 2006). *Poverty* in this study serves as a proxy for inequality. The principal argument of possible impact of poverty on GSP growth originates from the central concept of social divergence (Grafton et al. 2004). As higher level of poverty implies higher level of social segregation, the barriers to communication across the groups based on income, are most likely to have a negative impact on economic outcomes[x].

We include two interaction terms in our OLS estimation, generated with net migration rate and statistically significant diversity measures — *Language* and *Race*. The interaction terms are based on the data on net migration rate of foreign- born population for 1995-2000 or *Migration*, and is collected from the Bureau of Census (Perry and Schachter 2003). The recent studies on diversity and macroeconomic indicators, by Ottaviano and Peri (2005) and (Sparber 2006) treat the issues related to migration differently. Based on the rationale that the newly arriving immigrants have a tendency to move to places dominated by people with similar background, the authors construct new diversity measures based on 'shift-share' methodology (Card 2001). The constructed diversity measure is then used as an instrumental variable for respective diversity measures for each year. The use of this methodology is restricted for this analysis, as this investigation deals with cross-sectional data. For the interaction term, the net migration of the foreign-born population is considered more mobile than the native-born population. On a national level, 57.4 percent of foreign-born population[xi] reported living in a different residence in 2000 than in 1995, compared with 44.3 percent of the native-born population (Perry and Schachter 2003). Given the evidence that net migration rate is higher for the diverse populations, both linguistically and racially, the hypothesis is that the different rates of net migration of foreign-born populations are more likely to have differential impacts on GSP growth.

As the logarithm of initial income per capita is considered as a measure of distance from steady state, the hypothesized sign for the coefficient of initial income per capita is negative (Dowrik and Rogers 2002; Mankiw, Romer, and Weil 1992). In column (4), Table 2, the coefficient for initial GSP per capita provides strong support for growth specification with an expected negative sign and is significant at 10-percent level on the basis of a one-tail test. The coefficients for diversity measures in column (4) reveal similar results in comparison to estimates from the level framework in column 3. *Race* is not statistically significant at the usual significance level (p-value is 0.205), though the coefficient has the hypothesized negative sign. Both the interaction terms, although very small in value, are statistically significant at the 10-percent level. *Education* and *Poverty* are statistically significant at the 20-percent (with a p-value of 0.118) and 1-percent level respectively.

To check the robustness of our results with OLS, we use the general-to-specific modelling implemented in *PcGets* (Hendry 1995). In *PcGets* the estimation starts with a general unrestricted model and follows a multi-step procedure to select a preferred model where at each step there is an extensive set of diagnostic checks for consistency and statistical significance. Two principal requirements for *PcGets* modeling are 'congruency' and 'encompassing criterion'. Congruency relates to the matching of the model with the evidence in the data with respect to six main criteria (Hendry 1995:365): 'homoskedastic, and innovation errors; weakly exogenous conditioning variables for the parameters of interest; constant, invariant parameters of interest; theory-consistent, identifiable structures; data-admissible formulations on accurate observations, and encompassing of rival models'. The encompassing criterion ensures that no information is lost when the statistically insignificant variables are deleted when arriving at the specific model (Hendry 1995; Owen 2003).

We know that factors other than education, urbanization and poverty, are likely to influence the rate of GSP growth. Hence, the potential for omitted variables still exists with our model specification for OLS estimation. As inclusion of more regressors will lead to a loss of degrees of freedom, *PcGets* algorithm is specifically helpful for our estimation with only 48 data points. In comparison to the initial model, as all three additional regressors: poverty, and interaction terms for linguistic and racial diversity with net migration rate, are statistically significant, we choose the model represented in column 4, Table 2 as the *general* unrestricted model. The results in column (5) are the estimate of the *specific* model selected with application of the *PcGets* algorithm. At this stage, we verify the misspecification of functional form for *specific* model. We derive RESET test-statistics in two methods-with the fitted values, and with the square of endogenous variables. F-statistics are: 0.07 with a *p*-value of 0.9731 and 0.12, with a *p*-value of 0.9471 respectively. As the null hypothesis cannot be rejected with either of them, we conclude that there is no evidence of general functional form misspecification.

More significantly, column 5 in Table 2 indicates that the coefficients for both diversity measures — *Race*, and *Language* — and also the interaction terms with the net migration rate are statistically significant in the *specific* model. We test for statistical significance for interaction terms[xii], and conclude that at the average net migration rate of foreign population, racial diversity has a statistically significant negative effect on growth rate of GSP per capita; and linguistic diversity has a statistically significant positive effect on growth rate of GSP per capita.

The negative sign for *Race* supports our principal hypothesis and initial OLS estimates, i.e., racial barriers reduce GSP per capita growth. The positive sign for *Language*, however, needs clarification. This result implies that higher linguistic diversity increases GSP per capita growth, similar to the findings in Ottaviano and Peri (2005). The authors consider linguistic diversity as the most appropriate measure to capture the cultural divide in the US. We argue that differences in cultural or ethnic identity in a country like the US, after decades of intermarriage and the consequent multiethnic ancestry, may not necessarily be captured solely by the differences in mother tongue. Moreover, the census questionnaire for language data does not permit the determination of the primary or dominant language of the people who speak both English and another language[xiii]. Without assessing the information on the dominant language, the impact of linguistic diversity in knowledge spillovers will probably not be captured well with this dataset.

Assessing racial diversity, on the other hand, is less complicated as race is more easily identifiable. One of the criticisms against racial diversity as a proxy for cultural diversity argued by Ottaviano and Peri (2005) is the process of self-identification, i.e., people choose race(s) they most closely identify with. However, identification of ethnicity[xiv] and language ability is also self-assessed and can arguably be more subjective in nature for a resident in the US by comparison to race.

Section V: Robustness Checks

In this section, we present the regression results to check the robustness of our *specific* model derived in the previous section, and especially the effects of racial diversity with respect to: one, inclusion of other control regressors or the potential for omitted variables (Section 5.1); and two, other measures of diversity and different definition of English fluency (Section 5.2). We also check the robustness of the estimates to potential outliers in Section 5.3. To avoid biased estimation, instead of adding the additional control variable(s) to the specific model derived in the previous section, we add the additional variables to the unrestricted *general* model (column 4, Table 2). If the additional control variable is not included in the *specific* model, it is skipped for the next estimation, otherwise maintained in the subsequent estimations.

5.1 Robustness check with potential omitted variables

To test the robustness of the specific model derived in the previous section, first, we add two controls found to contribute to different economic variables across the US states: Manufacturing Employment and Research and Development (R&D) expenditure per capita. Manufacturing employment is included to account for the positive impact of agglomeration economies generated through industrialization[xv], as neither population density nor urbanization is significant in the previous estimations. It is measured as the percentage of population in non-farm employment in 2000, obtained from the State and Metropolitan Area Data Book (U.S. Department of Commerce 2006). Sedgley and Elmslie (2004) provide evidence that Research and Development (R&D) expenditure across the states in the US enhances the rate of innovative activity and, thereby, contributes to economic growth. Rauch (1993) also argues that it is likely to contribute through accumulation of human capital with the possibility of quicker knowledge diffusion. We, therefore, control for state R&D infrastructure by incorporating data on research and development expenditure per capita. Data on total expenditure on R&D (in million dollars) was obtained from collected from Science and Engineering Indicators (National Science Foundation 2003). Given the wide disparity across the states, we include natural logarithm of R&D per capita in the estimation.

We also examine another set of control variables that may contribute positively towards knowledge or information exchange by increasing the possibility of easier knowledge interactions. This set includes two mass communication variables that represent the percentage of households with telephone and percentage with access to internet services in 2000, and also a network variable — *Integration* — that represents the mean number of 'civic and social associations' per 1000 population, for the period 1977 to 1992. The data on mass communication variables are obtained from Statistical Abstracts of the United States: 2000, Bureau of Census. The data on integration is derived from Putnam's (2000) *Bowling Alone* dataset, compiled from *County Business Patterns, US Department of Commerce.* Mean number of civic and social organizations is considered as one of the measures of community organisational life and also used in the construction of comprehensive social capital index. Elaborating on the role of social capital, Putnam (2000:289) argues that 'communities that lack civic interconnections find it harder to share information'. *Integration* therefore, has been included as a control for social infrastructure facilitating exchange of information.

None of the potential omitted variables, manufacturing employment, R&D per capita, mass communication, and integration, is statistically significant. The *specific* model comprises the same regressors presented in column (5), Table 2, after the inclusion of all of the potential omitted variables in the general unrestricted model. The results of the estimations, therefore, are not reported separately to avoid repetition.

5.2 Robustness check: capturing diversity

In this section, the robustness check includes a set of additional controls with an attempt to capture different dimensions of social diversity in the US. First, the robustness of the statistically significant diversity measures (column 5, Table 2) is checked with an alternative measure of diversity; i.e. region-of-birth fractionalisation index. This index is measured from the groupings based on region of birth of foreign-born population. This diversity measure is likely to capture another dimension of social diversity or *cultural diversity* shaped by region of birth, and helps to verify the robustness of the estimates for the other two diversity measures. Based on census dataset, this diversity index has been constructed for six regions of birth for the foreign-born population: Europe, Asia, Africa, Oceania, Latin America and Northern America(Malone et al. 2003). The estimated coefficients confirm the prior result in column (5), Table 2, as the specific model retains the same set of variables after inclusion of cultural diversity index in the unrestricted general model.

To verify the impact of English fluency across the state population, we include *PIEF*, or percentage of the population with insufficient English fluency as a control regressor. This percentage provides information as to whether the communication barriers or transaction costs are relevant for GSP per capita growth. Using the Census dataset, we include percentage of population (5 years and above) who speak less than '*Very well'* that is a self-evaluation of English Fluency by non-native speakers. The non-native speakers can select in one of the following categories: *Very well, Well, Not well*, *or Not at all.* The choice of *PIEF*, however, is based on a English Language Proficiency Study by the Bureau of Census in 1982 (Kominski 1989). Proficiency tests were conducted on primarily non-native speakers along with a control group of native speakers. The test results showed that persons responding to the *Very well* criterion achieved the same level of passing as the control group and the persons responding to *Well* or worse had significantly higher level of failure. Percentage of state population who speak English less than *Very well*, therefore, has been included as a control to signify the cost of communication due the inability to interact in English.

Table 3 reports the regression results - column (1) for the general model and column (2) for the specific model, after inclusion of *PIEF* in the general model. *PIEF* is not only statistically significant at the 1-percent level of significance in the specific model, but with the inclusion of *PIEF Language* is no longer statistically significant in the specific model in column (2), Table 3. *Race*, on the other hand, is negative in both models, the general model with a *p*-value of 0.163 and with statistical significance at the 5-percent level in the specific model. The estimated impact of racial diversity on GSP per capita growth is slightly stronger as well compared with estimates in column (5), Table 2. In other words, when we control for low level of English fluency, or the inability to communicate effectively, the effect of racial diversity is stronger. If low level of English Fluency is positively correlated with the cost of communication across the population, the results support the view that the impact of racial diversity is negative when barriers of communication are higher due to low level of English fluency. The positive sign for *PIEF* is most likely to do with the more productive states like California and Florida, having the larger population of Spanish speakers or non-native workers. This matter is further investigated in the following sections.

To check on the robustness of *PIEF*, we construct another diversity index based on large linguistic groups. Based on the detailed list of language spoken at home for the state population 5 years and over (U.S. Department of Commerce 2000), linguistic groupings comprise languages with common roots:

i. Anglo-Saxon or Germanic (English, German, Yiddish, Scandinavian, Other west-Germanic languages);
ii. Neo-Latin (Italian, Portuguese or Portuguese Creole, French or French Creole, Spanish or Spanish Creole);
iii. Slavic (Russian, Polish, Serbo-Croatian, Other Slavic languages);
iv. Indo-Iranian (Persian, Gujarathi, Hindi, Urdu, other Indic languages);
v. Greek (Greek, Armenian);
vi. Asian (Chinese, Japanese, Korean, Mon-Khmer, Cambodian, Miao, Thai, Laotian, Tagalog, Vietnamese and other Asian languages);
vii. Native American (Navajo, other Native American languages);
viii. Other languages (Arabic, Hebrew, African languages and others)

Identification of groups based on common language roots, to some extent, captures the origin of cultural similarities[xvi]; and helps us to verify the impact of *cultural diversity* from a different perspective by comparison with the birth diversity index of the foreign-born population. This measure of linguistic diversity, however, proves to be statistically insignificant at the 10 per cent level of significance because the specific model reported in column 2, Table 3, remains unchanged after inclusion in the general model. Consequently, we explore the influence of *PIEF* with another alternative indicator for low level of English fluency, namely *PLIP* or percentage of linguistically isolated population. Based on the data provided on *linguistic isolation*[xvii], we measure the percentage of linguistically isolated population. Column (3), Table 3 provides the estimates for the general model. The coefficient for *PLIP* is significant at the 5-percent level. For diversity measures, the results are qualitatively the same as Column (1), Table 3 with the hypothesized negative sign for *Language* and *Race*. None of them, however, appears to be statistically significant at the conventional level of significance. The estimates for the specific model are reported in Column (4), Table 3. The coefficient for racial diversity is statistically significant at the 5-percent level under the specific model, and the estimates are very similar to the specific model reported in column (2), Table 3.

Although we do not have the data at the state level on percentage of linguistically isolated population who speak Spanish at home, a linguistically isolated population is more likely to be dominated by the Spanish speaking population. In 1990, 54.44 percent of the population who does not speak English at home are Spanish speakers. Spanish continued to be the most widely spoken non-English language in 2000, as 59.85 percent of non-English speakers were Spanish. In relation to the linguistically isolated population, also notable is the increase in the number of Chinese speaker in recent years. In the 1990 census Chinese was ranked fifth among the most frequently spoken non-English language. Although Chinese has been ranked as the second most widely spoken non-English language, only 4.31 percent of non-English speakers are reported to speak Chinese at home in 2000. To check the robustness of linguistic isolation, we therefore, include a dummy variable, *Latin Dummy* in the general model. *Latin Dummy* accounts for the effects of proximity to Mexico and other Spanish-speaking countries of Latin America. *Latin Dummy* is not statistically significant in the general model and the estimates for the specific model remain the same as in column (4), Table 3.

In summary, we conclude that the estimated impact of linguistic diversity is sensitive to the inclusion to variables representing ability to interact, i.e., *PIEF* or *PLIP* or cost of communication. The estimated impact of racial diversity is, however, robust as both the specific models in column (2) and (4) in Table 3 retain the hypothesized negative value for *Race*, which in both cases is statistically significant at the 5-percent level.

5.3 Robustness check: biased estimation

In this section, we address two issues: detection of influential observations or outliers and heteroskedasticity. First, we rerun the final specific model in column (4), Table 3 to identify potential outliers or influential observation. The estimation with Robust Regression Analysis is presented in Column (1), Table 4. Although there is a very small reduction in value in all of the coefficients and a decrease in overall goodness of fit, the results are qualitatively unchanged. The main variable of interest, racial diversity, has the hypothesized negative sign and is statistically significant at the 10-percent level under robust regression estimation.

To address the potential outlier influence, we re-run the regression of the final specific model, adopting median regression or Least Absolute Deviation (LAD) estimation method. The results are reported in column (2), Table 4. In comparison with the OLS estimates and robust regression estimation, the results under LAD are qualitatively same. The value of the coefficients are slightly increased for *Education, Race and Poverty* while the t-ratio is increased for all the regressors except the constant and *lnGSP99* compared to the estimates in column (4), Table 3. Despite the significant level of subjectivity attached to the concept of deleting outliers, especially in the case of a small sample size (Gomanee, Girma, and Morrissey 2002), we re-run the final specific model without two potential outliers, i.e., Oregon and Kentucky. The absolute values of the standardized residuals for Oregon and Kentucky are 2.4979 and 2.3302 respectively. We stress that none of the observations has been identified as a gross outlier by Cook's Distance measure and hence are not excluded under robust regression[xviii]. We present the regression results in column (3), Table 4. The estimates are qualitatively similar for all regressors. The main variable of interest, racial diversity, has retained the hypothesized negative sign with an increase in the level of significance in comparison with the robust regression estimate in column (1), Table 4.

Lastly, column (4), Table 4 provides the results of the final specific model with robust standard errors. No major differences emerge with the estimation of robust standard error, as all coefficients are statistically significant using the heteroskedasticity-robust t-statistics. We then derive the Breusch-Pagan (BP) Test Statistic for detecting heteroskedasticity. For the final specific model, BP=6.13 (*p*-value=0.5248) and the null hypothesis of homoskedasticity cannot be rejected as $\chi^2(7) = 12.02$ at the 10-percent significance level. The reported White-Hetero Test Statistic for the specific model in column (4), Table 3 is 0.6829 (*p*-value=0.7693), which also provides support for homoskedasticity in our earlier estimation.

Section VI: Evaluating impact of endogeniety

In this section we verify potential endogeniety of three variables with: *Race*, *Education* and *Poverty.* Two methods are employed for checking endogeniety: deriving the *Hausman* test to check for consistency of OLS estimate of the specific model in column (4), Table 3; and testing the significance of predicted error in the reduced-form equation.

6.1 Instrument for *Race*

Most of the cross-country studies treated fractionalization measures as. Given the arguments of reverse causation in more productive states like New York or California, and also because of the nature of domestic migration of the foreign-born population in the US, potential endogeniety for *Race* needs to be addressed.

The lagged value of *Race*, i.e., racial diversity in 1990 is used as an instrument, because using measures from 1990 can, to some extent, reduce reverse causality issues. High levels of GSP or income per capita in 2000 may have spurred migration to cities in 2000, but racial diversity a decade back is more likely to causally impact GSP per capita in 2000, rather than GSP per capita in 2000 impacting racial diversity in 1990. Racial categories, however, are defined differently in 1990, being revised by the Office of Management and Budget (OMB) in 1997 (Grieco and Cassidy 2001). The 1990 Census did not include Native Hawaiian and other Pacific Islanders as a separate category. This particular category is less likely to have an influence on the measurement of diversity because the population in this category constitutes only 0.1 percent of total population in 2000.

The second instrumental variable we consider is the Comprehensive Social Capital Index (*SCI*)[xix] at the state level developed by Putnam (2000). This state-level measure for participation was compiled from a set of indicators representing different facets of social and community life, which contribute to building social trust and connectedness. The rationale for this instrument can be derived on the grounds that literature provides evidence that higher levels of racial or ethnic diversity influence the level of trust, and, thereby, influence the nature and frequency of participation in associational activities for collective problem solving. A more recent study by Rupasingha et al. (2006) measures social capital index at the county level in the US. The robustness of the negative fractionalization coefficient for the 1980-90 period is retained when the authors develop separate sets of social capital indexes based on associational density of two types of organizations[xx]—'Olson-type', rent-seeking organizations and 'Putnam-type', organizations promoting trust and social cooperation.

Apart from evidence of correlation between different measures of social capital and racial fragmentation, a comprehensive social capital index also accounts for historical factors like slavery that are not considered as a separate instrument for *Race*, but are likely to influence racial diversity in the US (Sparber 2006a). Putnam (2000) argues that the lowest level of social capital index in the southern states of the US is strongly associated with the history of plantation slavery in that region.

To check the validity of our assumption of racial diversity being endogenous, first, we derive the *Hausman* test statistics reported in Table 5. For lagged diversity as an instrument for

Race, the *Hausman* test statistic is 0.13 with a *p*-value of 1.000. In column (2), Table 5, for the social capital index, the test statistic is 0.74 with a p-value of 0.9935. For both the instruments, the null hypothesis that the OLS estimate is consistent cannot be rejected. *Hausman* test statistics, in these cases, provide evidence of consistency of the OLS estimate and consider instrumental variable estimation (IVE) as less efficient.

Racial diversity proves to be potentially endogenous, when we test for statistical significance of predicted error of reduced form equation:

$$Z_i = \pi_0 + \pi_1 X_i + \pi_2 IV_i + v_i \text{ — (6.1)}$$

In equation (6.1), Z_i is the respective endogenous variable, X_i includes the exogenous variables in the specific model in column (4), Table 3, and IV_i represents instrumental variable used. Table 5, columns (1)-(2), report the estimation of the reduced form equation for racial diversity with two instrumental variables. In both cases $\hat{\pi}_2$ is statistically significant at the 1-percent level with respective hypothesized signs. Thus both lagged diversity and the social capital index have a strong association with *Race*.

To assess whether instruments are uncorrelated with the error term in the structural equation, the predicted residual, $\hat{v}_1$ from the reduced form equation was added to structural equation, i.e.,

$$\frac{1}{T}\ln\left(\frac{GSP_{iT}}{GSP_{it}}\right) = \pi + \lambda X_i + \alpha Z_i + \theta \hat{v}_1 + \varepsilon_i \quad \text{— (6.2)}$$

We then estimate equation (6.2) to test the null hypothesis, $H_0: \theta = 0$. The estimated $\hat{\theta}$ =1.213 with a t-statistics of 1.43. The null hypothesis can be rejected at the 20 percent significance level. This is moderate evidence of correlation of ε and $\hat{v}_1$. The same steps are followed to test the statistical significance of the predicted error from the reduced form equation with *SCI.* In this case $\hat{\theta}$ = 0.246 with a t-statistic of 1.81. The null hypothesis is rejected at the 10 percent significance level.

The statistical significance of the predicted errors from the reduced form equation for both the instruments, thus, provides support of instrumental variable estimation to address the endogeniety of *Race*.

6.2 Instrument for *Education*

Following two recent studies by (Basher and Lagerlöf) 2006) and Shapiro (2006), we explore the possible endogeniety of the education variable. (Basher and Lagerlöf) (2006) argue that geography, i.e., temperature, average rainfall etc., mattered for the choice of early European settlers. As the coastal states of the US are still the most densely populated areas and have a high-skill concentration, the authors argue against the emphasis on institutions (Acemoglu 2002) as the intermediate factor between geography and economic outcomes. Drawing on the strong relation between agglomeration variables and human capital, the authors, therefore, use a set of instrumental variables for education: population density in 1900, sex-ratio in 1900, fraction of slaves in the population in 1850, average temperature, average precipitation, average rainy days, Atlantic dummy and Great Lakes Dummy. Shapiro (2006), however, uses a significantly different set of instruments, focusing on the relation between institutions and human capital accumulation- land grant institutions (Moretti 2004 cited in Shapiro 2006) and compulsory schooling laws (Acemoglu and Angrist 2000 cited in Shapiro 2006). The two instrumental variables for education are: percentage of slave population in 1850 or *Slavery* and population density in 1850 or *PopDensity00*, from the *Geospatial and Statistical Data Center* at the University of Virginia Library[xxi].

Putnam argues that social segregation[xxii] was institutionalised under slavery. The lack of 'weak ties' under slavery can also be linked to our principal hypothesis of social divergence. Basher and Lagerlof (2006) also argue that even after abolition of laws that did not allow the slaves to be educated, the effects seemed to have lingered on for a significant time period. Column (3), Table 5 reports the reduced form equation when education is instrumented with percentage of slave population. The coefficient, $\hat{p}_2$, does not have the hypothesized negative sign and is not statistically significant even at the 20 percent significance level. The percentage of slave population, therefore, is a weak instrument for education. The estimated $\hat{q}$ in equation (6.2) is -0.1379 with a t-statistics of 0.41. The *p*-value is 0.683 and the null hypothesis cannot be rejected at any conventional significance level. As there is no evidence of correlation of e and $\hat{v}$, we cannot establish that education is potentially endogenous. The null hypothesis for the Hausman statistic, with a *p*-value of 0.3927, cannot be rejected and thus implies consistency of OLS estimates in comparison to instrumental variable estimation.

The use of population density in 1900 as an instrument for *Education*, addresses the reverse causality issue and also incorporates the principal argument of agglomeration economies that population density enhances exchanges of ideas and knowledge and thus contributes to accumulation of human capital. Column 4, Table 5, reports the estimation for the reduced form equation when education is instrumented with log of *PopDensity00*. The coefficient for the instrument, despite having the hypothesized positive sign, is not statistically significant. Population density in 1900 is, therefore a weak instrument for education. The same steps are followed to test the statistical significance of the predicted error from the reduced form equation with population density in 1900. In this case $\hat{q}$ = 0.2937 with a t-statistics of 0.95. As the null hypothesis, H_0:q =0 cannot be rejected at the conventional significance level, endogeniety of education cannot be established. Hausman test statistics for both instruments in column (3)-(4), Table 5 also support this conclusion. As both the tests fail to prove endogeniety for education, we skipped education as a potential endogenous variable for instrumental variable estimation in the next section.

6.3 Instrument for *Poverty*

The empirical evidences in a study by Iceland et al. (2005), however, may question the validity of social divergence hypothesis for US economy on grounds of reverse causality. Iceland et al. (2005) investigate the effect of macroeconomic performance on two measures of poverty i.e. absolute and relative poverty during 1980s and 1990s. Their study results suggest that higher per capita GSP contributes to lower absolute poverty through increasing work hours and high low-end wage levels. To check for potential endogeniety associated with poverty, we use two instruments: social capital index and percentage of slave population in 1850.

Column 5, Table 5, reports the estimation for the reduced form equation when poverty is instrumented with the percentage of slave population in 1850. The coefficient for *Slavery*, has the hypothesized positive sign, but can be considered as a weak instrument for poverty with a p-value of 0.168. The predicted error from the reduced form equation with the percentage of slave population, however, proves to be statistically significant with a t-statistic of 1.63. The null hypothesis is rejected at the 15-percent level and thus provides limited evidence of endogeniety. The Hausman test statistic in column (5), Table 5, is 8.95 with a p-value of 0.1766. The null hypothesis that the OLS estimate is consistent is rejected implying that IVE is appropriate. Both the tests indicate the potential endogeniety of *Poverty.*

A similar conclusion can be drawn using *SCI* as an instrument for poverty. As poverty is considered as an indicator of inequality, the hypothesized sign for the social capital index is negative. Column 6, Table 5 reports the reduced form equation, when *Poverty* is instrumented with social capital index. In column (6), Table 5, the coefficient of the instrumental variable has a negative sign (with a *p*-value of 0.102). We then estimate the coefficient for predicted error, $\hat{v}$ and estimate equation (6.2); and find that $\hat{q}$ is 1.1921 with a t-statistic of 2.11. This establishes the evidence of correlation between e and $\hat{v}$ and the null hypothesis is rejected at the 5-percent significance level. In column (6), Table 5, for *SCI*, the Hausman test statistic is 6.97 with a *p*-value of 0.3241. As both the tests for both instruments indicate, to some extent, a potential endogeniety bias, *Poverty* is treated as an endogenous variable for the next set of estimations.

6.4 Two-stage least square estimation/Instrumental Variable Estimation

Table 6 presents results obtained using IVE in which fractionalisation measures and poverty are treated as potentially endogenous. To check for the validity of instruments used, Sargan test statistics are reported for each model specification. In addition to instruments discussed in the previous section, we use a set of geography variables as instruments in relation to the hypothesis that geography influences the productivity growth through climate, endowment of natural resources and disease. In relation to differences of GSP growth rate across the states in the US, Basher and Lagerlöf (2006) argue that geography contributed to the choice of early settlement by Europeans and contributes to GSP growth as a result of human capital accumulation in densely populated areas. Three geography variables are added as instrumental variables: average mean temperature (*MeanTemp)*, average precipitation (*AvegPrecip*) and average number of rainy days per year (*AvegRain*). We retrieve the data from the dataset provided by Basher and Lagerlöf(2006) for public use[xxiii].

Given the significance of racial diversity, at first we address the potential endogeniety of *Race* only. Column (1), Table 6 reports the IVE results for the final specific model in column (4), Table 3, using the extended set of instrumental variables. Sargan test statistic i.e. c^2 (5) =3.6703 with the reported p-value of 0.5978, supports the validity of the set of instruments used. A very high partial R^2 for first stage regression also signifies the strong association between race and the set of instruments. The result with IVE is consistent with OLS estimate presented in column (4), Table 3, namely, that higher racial diversity reduces GSP growth. The rest of the regressors are statistically significant at the conventional significance levels. Column (2), Table 6 reports the IVE results for the same specific model when both race and poverty are considered as endogenous variables. Partial R^2 from the first stage regression indicates the high correlation between the set of instruments and each of the endogenous variables. The value of the Sargan test statistic, with a *p*-value of 0.5046 supports the validity of the instruments used. The coefficient for *Race* retains the hypothesized negative sign, with only a very limited statistical significance with a p-value of 0.297. The results are similar for rest of the variables as in column (4), Table 3, apart from *Poverty*. With a *p*-value of 0.212, the coefficient for *Poverty*, retains the hypothesized negative sign.

Columns (3)-(4) in Table 6 report the final models obtained by applying the *Getsive* method, which follows similar general-to-specific selection criteria as in the *PcGets* algorithm, but adopts IVE instead of OLS. Sargan test statistics confirm the validity of instruments under both specifications. Beginning with the general model specified in column (3), Table 3, the results obtained through the *Getsive* algorithm are reported in column (3), Table 6. In comparison to the specific model obtained through OLS in column (4), Table 3, there is a slight increase in estimated impact of *Race* as well as in statistically significance, i.e., *Race* is significant at the 1 percent level. Although the estimated impact is very small, racial diversity and linguistic diversity remain significant through interaction terms as well.

To signify the robustness of the results for racial diversity, column (4), Table 6 reports the final model with the *Getsive* algorithm, applied to the first general model in column (4), Table 6. This specification excludes any of the variables representing ability to interact, PIEF and PILP. To check endogeniety of linguistic diversity, we compare the Hausman test statistics for IVE for the specific model with two specifications. If we consider linguistic diversity as a potential endogenous variable, the null hypothesis that the OLS estimate is consistent, is rejected even at the 1-percent significance level[xxiv]. This implies OLS estimates are inconsistent in comparison with IV estimate with three potential endogenous variables, including linguistic diversity. On the other hand, when only racial diversity and poverty are considered as endogenous variables, OLS estimates are consistent when compared to IV estimate where linguistic diversity is not considered endogenous[xxv].

As *Getsive* applies IVE instead of OLS, linguistic diversity is considered an endogenous variable for results in column (4), Table 6. The results are similar for all the variables, except poverty, in comparison with the specific model with OLS in column (5), Table 6. With the exclusion of PILP, *Language* is statistically significant at the 1-percent level in the specific model.

Column (5), Table 6 reports the IVE results of the final specific model after the inclusion of the interaction term with diversity measures and the geography variables as instruments. The objective is to account for the endogeniety of the interaction terms as regressors, i.e., *Language*Migration* and *Race*Migration.* High values of R^2 from the first stage regression specify that there is a strong correlation between the set of instruments and the corresponding endogenous variable that is supported by the Sargan test statistics. The results with IV estimation in column (5), Table 6 are very similar to the OLS estimates in column (4), Table 3. The estimated impact of all the variables has, however, increased with the IV estimation. In sum, the inclusion of interaction terms in the instrument set ensures all variables are statistically significant at the conventional significance level. Most importantly, the estimated impact of *Race* has increased and is statistically significant at the 10-percent level of significance. The coefficient for *Poverty* has retained the negative sign and is statistically significant at the 10-percent level.

To check the robustness of the racial diversity measure, we apply the *Getsive* method for the general model in column (3), Table 3 with interaction terms in the instrument set. The estimates are reported in column (6), Table 6. Sargan test statistics confirm the validity of instruments and the p-value is much higher than the IVE of the specific model in column (5), Table 6. In comparison with the *Getsive* results in column (3), Table 6 the variables selected with the extended set of instruments, are identical apart from poverty. Column (6), Table 6 reports that poverty is statistically significant at the 1-percent level. The variable of interest, *Race* retains its negative sign with a slightly higher value compared to OLS estimation and is statistically significant at the 1-percent level.

Section VII: Discussion

Our results indicate that the potential for the productivity-augmenting role of diversity is likely, only to be fully realized with the implementation of public policies aimed at reducing the social barriers to communication. We briefly explore the policy implications for two contentious areas in the present context: i immigration policies, and ii provision of public goods like education or health.

Immigration policies in settler societies aim to ensure economic assimilation of skilled migrants. Economic assimilation implies labour market assimilation[xxvi]. This group of migrants is targeted because of the potential for economic gain with their complementary knowledge, skills or experiences. Our results are not contrary to this policy but it does emphasize that social isolation of *'culturally distant'* immigrant communities can lead to underutilization of their skills and knowledge. Despite working as a springboard for undereducated and newly arrived immigrants, ethnic enclaves like 'Chinatown' or 'Little Italy' can retard the process of economic assimilation. For example, 43 percent of surveyed business entrepreneurs in British Columbia from three major immigrant communities, Hong Kong, Korea and China, identify that use of mother tongue as the business language is one of the main barrier for their success (Ley 2005). Although the success of Korean business migrants in Vancouver can significantly be attributed to their education level, their communication across the networks or decision to break the ethnic ties, contributed positively towards their commercial success. We would argue, on the basis of results for the US, that public policies that include English language training programs, which differentially targets foreign-born workers could significantly help economic assimilation of disadvantaged immigrant groups. For instance, a recent study by (Chiswick, Lee, and Miller) 2006) using Australian data provides evidence that language proficiency of immigrants is closely related with linguistic distance of their mother tongue. Differential rates of language acquisition skills due to linguistic distance, defined as the extent of difference between the origin and destination language, needs to be considered as part of the existing government-sponsored or subsidized language-training program.

Education policy can also be used as a tool to reduce social barriers to communication. Despite huge gains over the past decades social cleavages remain in the US, in terms of racial divisions. In some western liberal democracies social segregation between natives and ‘culturally distant’ immigrant communities is becoming even more evident as the immigrant groups become larger. As there are studies that indicate that there is a positive association between education and associational activities and collective action (Putnam 2000; Rupasingha, Goetz, and Freshwater 2006; Sacerdote and Glaeser 2001), our results would suggest there is a need for public polices to promote social integration as part of broader education policy. This would involve the integrationist concept of ‘two-way traffic’; so that a curriculum on intercultural education highlights the contribution of ethnic minorities to the wider society as well as the core values of the dominant culture.

Section VIII: Concluding remarks

We investigate the empirical evidence of social divergence hypothesis that social barriers to communication have negative impacts on aggregate economic performance. Employing the US Census 2000 dataset for 48 continental states, we find that the estimated impact of racial diversity, obtained from ordinary least squares estimates as well as with instrumental variables, is negative on GSP per capita growth, when the barriers to communication are higher. The results provide fresh insights into overcoming the barriers to communication across social groups as a public policy tool that generates positive economic effects.

Using US Census dataset we are able to show that linguistic isolation and racial measures of social fractionalization contribute to lower stale-level economic performance after controlling for other variables. The robustness of the estimation results provides strong support for the theoretical model by (Grafton, Kompas, and Owen) 2007) on economic effects of social barriers to communication. Namely, the results support the view that the impact of racial diversity is negative when barriers to communications are higher due to low level of English fluency. These findings are supported with instrumental variables estimation where an extensive set of instruments is employed for two potential endogenous variables: racial diversity and poverty. Overall, the empirical results provide strong support for the economic significance of establishing ‘weak ties’ across social groups in the US that will help overcome barriers to communication that may be defined by race, ethnicity and language.

References

Acemoglu, D. "Reversal of Fortune: Geography and Institutions in the Making of the Modern World Income Distribution." *Quarterly Journal of Economics* 117, 2002.1231-1294.

Alesina, A., R. Baqir, and W. Easterly. "Public Goods and Ethnic Divisions." *Quarterly Journal of Economics* 114 (4), 1999.1243-1284.

Alesina, A., A. Devleeschauwer, W. Easterly, S. Kurlat, and R. Wacziarg. "Fractionalization." *Journal of Economic Growth* 8 (2), 2003.155-194.

Alesina, A., and Ferrara. "Ethnic Diversity and Economic Performance." *Journal of Economic Literature* 43 (3), 2005.762-800.

Alesina, A., and E. L. Ferrara. "Participation in Heterogeneous Communities." *The Quarterly Journal of Economics* August, 2000.847-904.

Alesina, A., and E. L. Ferrera. "Who Trusts Others?" *Journal of Public Economics* 85, 2002.207-234.

Antecol, H., D. A. Cobb-Clark, and S. J. Trejo. "Immigration Policy and the Skills of Immigrants to Australia, Canada, and the United States." *The Journal of Human Resources* 38 (1), 2003.192-218.

Axelrod, R. Agent-Based Modelling as a Bridge between Disciplines. In *Handbook of Computational Economics, Volume II:Agent-Based Computational Economics*, edited by L. Tesfatsion and K. L. Judd. Amsterdam: Elsvier/North-Holland. 2006.

Basher, S. A., and N.-P. Lagerlöf. Geography, Population Density and Per-capita Income Gaps across US States and Canadian Provinces. Munich: University Library of Munich. 2006.

Beeson, P. "Total Factor Productivity Growth and Agglomeration Economies in Manufacturing, 1959-1973." *Journal of Regional Science* 27 (2), 1987.183-199.

Beiser, M., and F. Hou. "Gender differences in language acquisition and employment consequences among Southeast Asian refugees in Canada." *Canadian Public Policy-Analyse De Politiques* 26 (3), 2000.311-330.

Brittingham, A., and G. P. de la Cruz. Ancestry :2000 edited by U. S. D. o. C. Economics and Statistics Administration: U.S. Census Bureau. 2004.

Chiswick, B. R., Y. L. Lee, and P. W. Miller. "Immigrants' Language Skills and Visa Category." *International Migration Review* 40 (2), 2006.419-450.

Ciccone, A., and R. E. Hall. "Productivity and the density of economic activity." *American Economic Review* 86 (1), 1996.54-70.

Collier, P., and J. W. Gunning. "Explaining African Economic Performance." *Journal of Economic Literature* 37 (1), 1999.64-111.

Dowrik, S., and M. Rogers. "Classical and Technological Convergence." *Oxford Economic Papers* 54 (3), 2002.369-85.

Easterly, W., and R. Levine. "Africa's Growth Tragedy: Policies and Economic Divisions." *The Quarterly Journal of Economics* 112 (4), 1997.1203-1250.

Florida, R. "Bohemia and Economic Geography." *Journal of Economic Geography* 2 (1), 2002.55-71.

Fu, S. "Smart Cafe Cities: Testing Human Capital Externalities " *Journal of Urban Economics* 61, 2007.86-111.

Glaeser, E. "Are Cities Dying?" *Journal of Economic Perspectives* 12 (2), 1998.139-160.

Glaeser, E. L., H. D. Kallal, J. A. Scheinkman, and A. Shleifer. "Growth in Cities." *Journal of Political Economy* 100 (6), 1992.1126-1152.

Goldin, C., and L. Katz. "Human Capital and Social Capital: The Rise of Secondary School in America, 1910 to 1940." *Journal of Interdisciplinary History* 29, 1999.683-723.

Gomanee, K., S. Girma, and O. Morrissey. Aid and Growth in Sub-Saharan Africa: Accounting for Transmission Mechanisms. Nottingham: University of Nottingham. 2002.

Grafton, R. Q., S. Knowles, and D. P. Owen. "Total Factor Productivity, Per Capita Income and Social Divergence." *The Economic Record* 80 (250), 2004.302-313.

Grafton, R. Q., T. Kompas, and D. P. Owen. "Bridging the Barriers: Knowledge Connections, Productivity and Capital Accumulation." *Journal of Productivity Analysis* 27 (3), 2007.219-231.

Granovetter, M. S. "The Strength of Weak Ties." *The American Journal of Sociology* 78 (6), 1973.1360-1380.

Grieco, E. M., and R. C. Cassidy. Overview of Race and Hispanic Origin:2000, edited by U. S. D. o. C. Economics and Statistics Administration: U.S. Census Bureau. 2001.

Hall, R. E., and C. I. Jones. "Why Do Some Countries Produce so Much More Output Per Worker than Others?" *The Quarterly Journal of Economics* 114 (1), 1999.83-116.

Hendricks, L. Why does educational attainment differ across U.S. states? Munich: Department of Economics, University of Munich. 2004.

Hendry, D. F. *Dynamic Econometrics*. Oxford: Oxford university press. 1995.

Higgins, M. J., D. Levy, and A. T. Young. "Growth and convergence across the United States: Evidence from county-level data." *Review of Economics and Statistics* 88 (4), 2006.671-681.

Hum, D., and W. Simpson. "Closing the Wage Gap: Economic Assimilation of Canadian Immigrants Reconsidered." *Journal of International Migration and Integration*, 2001.

Iceland, J., L. Kenworthy, and M. Scopilliti. Macroeconomic Performance and Poverty in the 1980s and 1990s: A State-Level Analysis. Madison:

Institute for Research on Poverty, University of Wisconsin-Madison. 2005.

Knack, S., and P. Keefer. "Does Social Capital Have an Economic Payoff? A Cross-Country Investigation." *The Quarterly Journal of Economics* 112 (4), 1997.1251-1288.

Kominski, R. How Good is "How Well"?: An Examination of the Census English-Speaking Ability Question. In *Annual Meeting of the American Statistical Association*. Washington, DC. 1989.

Leigh, A. "Trust, Inequality and Ethnic Heterogeneity." *Economic Record* 82 (258), 2006.268-280.

Ley, D. Indicators of Entreprenurial Success among Business Immigrants in Canada. In *Research on Immigration and Integration in the Metropolis*. Vancouver Vancouver Centre of Excellence. 2005.

Malone, N., K. F. Baluja, J. M. Costanzo, and C. J. Davis. The Foreign-Born Population:2000, edited by U. S. D. o. C. Economics and Statistics Administration: U.S. Census Bureau. 2003.

Mankiw, N. G., D. Romer, and D. N. Weil. "A Contribution to the Empirics of Economic Growth." *The Quarterly Journal of Economics* 107 (2), 1992.407-437.

Meng, X., and R. G. Gegory. "Internarriage and Economic Assimilation of Immigrants " *Journal of Labor Economics* 23 (135-175), 2005.

National Science Foundation. Science and Engineering Indicators. 2003.

Ottaviano, G. I. P., and G. Peri. "Cities and Culture." *Journal of Urban Economics* 58, 2005.304-337.

Owen, D. P. "General-to-specific Modelling Using PcGets." *Journal of Economic Surveys* 17 (4), 2003.609-628.

Perry, M. J., and J. P. Schachter. Migration of Natives and the Foreign-Born: 1995 to 2000, edited by U. S. D. o. C. Economics and Statistics Administration: U.S. Census Bureau. 2003.

Persson, J., and B. Malmberg. Human Capital, Demographics and Growth Across US states. In *Seminar Paper 619*. Institute for International Economic studies, Stockholm University 1996.

Poortinga, W. "Social relations or social capital? Individual and community health effects of bonding social capital." *Social Science & Medicine* 63 (1), 2006.255-270.

Poterba, J. "Demographic Structure and the Political Economy of Public Education." *Journal of Policy Analysis and Management* 16 (1), 1997.48-66.

Putnam, R. D. *Bowling Alone: The Collapse and Revival of American Community*. New York: Simon and Schuster. 2000.

Rauch, J. E. "Productivity Gains from Geographic Concentration of Human Capital: Evidence from Cities." *Journal of Urban Economics* 34, 1993.380-400.

Reynal-Queral, M. "Ethnicity, Political systems and civil wars." *The Journal of Conflict Resolution* 46, 2002.29-54.

Rigby, D. L., and R. Essletzbichler. "Agglomeration Economies and Productivity Differences in US Cities." *Journal of Economic Geography* 2 (4), 2002.407-432.

Rupasingha, A., S. J. Goetz, and D. Freshwater. "The production of social capital in US counties." *The Journal of Socio-Economics* 35, 2006.83-101.

Sacerdote, B., and E. L. Glaeser. Education and Religion. Cambridge, Massachusetts: Working Paper 8080, National Bureau of Economic Research (NBER). 2001.

Sedgley, N., and B. Elmslie. "The Geographic Concentration of Knowledge: Scale, Agglomeration, and Congestion in Innovation Across U.S. States." *International Regional Science Review* 27 (2), 2004.117-137.

Segal, D. "Are there Returns to Scale in City Size?" *The Review of Economics and Statistics* 58 (3), 1976.339-350.

Shapiro, J. M. "Smart Cities: Quality of Life, Productivity and the Growth Effects of Human Capital." *The Review of Economics and Statistics* 88 (2), 2006.324-335.

Shin, H. B., and R. Bruno. Language Use and English-Speaking Ability:2000, edited by U. S. D. o. C. Economics and Statistics Administration: US Census Bureau. 2003.

Simon, C. J., and C. Nardinelli. "Human Capital and Rise of American Cities, 1900-1990." *Regional Science and Urban Economics* 32, 2002.59-96.

Sparber, C. Racial Diversity and Macroeconomic Labor Productivity in US Industries:1980-2000. Colgate University, New York. 2006.

———. Racial Diversity and Macroeconomic Productivity Across US States and Cities. Davis, California. 2006.

Stolzenberg, R. M., and M. Tienda. "English Proficiency, Education and the Conditional Economic Assimilation of Hispanic and Asian Origin Man." *Social Science Research* 26 (1), 1997.25-51.

Temple, J., and P. A. Johnson. "Social Capability and Economic Growth." *The Quarterly Journal of Economics* 113 (3), 1998.965-990.

U.S. Department of Commerce. Census of Population and Housing, Census 2000 Summary File 3, U.S. Census Bureau. 2000.

———. Statistical Abstracts of United States, U.S. Census Bureau. 2003.

———. State and Metropolitan Area Data Book, U.S. Census Bureau. 2006.

———. State Data Tables, State and Metropolitan Area Data Book, U.S. Census Bureau. 2006.

US Department of Commerce. Gross State Product: Bureau of Economic Analysis, Regional Economic Accounts. 2004.

Woodridge, J. M. *Introductory Econometrics: A Modern Approach*. Ohio: Thompson. 2003.

Zak, P. J., and S. Knack. "Trust and Growth." *The Economic Journal* 111 (April), 2001.295-321.

APPENDIX A: SOURCES AND DEFINITIONS

lnGSP: Gross State Product per capita (in natural logs), Source: US Bureau of Economic Analysis, 1999-2000

Race, Language, Religion: Fractionalisation indexes for racial, linguistic and religious groups. Source: US Bureau of Census (Beiser and Hou 2000)

Education: Percentage of population with a bachelor degree or higher, Source: US Bureau of Census (2000)

Urbanization: Percentage of state population living in metropolitan areas, Source: US Bureau of Census (Alesina and Ferrara)

Age: Percentage of state population in within the range of 20-64 years, Source: US Bureau of Census (2000)

PopDensity: Average number of inhabitants per square mile of land area, Source: US Bureau of Census (Alesina and Ferrara)

Poverty: Percentage of families below the poverty line in the immediate past (last 12 months), Source: State and Metropolitan Area Data Book, US Bureau of Census (Axelrod)

Migration: Net migration rate of foreign-born population, Source: US Bureau of Census (1995-2000),

PIEF: Percentage of population (Poortinga) with insufficient English Fluency, Source: US Bureau of Census (2000)

PLIP: Percentage of linguistically isolated population (14 years and above), Source: US Bureau of Census (2000)

Lagged diversity: Fractionalisation index for racial groups in 1990, Source: US Bureau of Census (Antecol, Cobb-Clark, and Trejo)

SCI: Comprehensive Social Capital Index at State level, Putnam (2000)

Slavery: Percentage of slave population in 1850, Source: Geospatial and Statistical Data Center, University of Virginia Library

PopDensity00: Population density in 1850, Geospatial and Statistical Data Center, University of Virginia Library

Table 1: Summary statistics for key variables

Variable	Period	N	Mean	Standard Deviation	Minimum	Maximum
lnGSP	2000	48	4.5101	0.0777	4.3530	4.7328
lnGSP99	1999	48	4.5005	0.0737	4.3586	4.7246
GSP growth	1999-2000	48	0.0095	0.0105	-0.0175	0.0306
Race	2000	48	0.2864	0.1407	0.0402	0.5623
Language	2000	48	0.2109	0.1285	0.0528	0.5578
Religion	2000	48	0.9462	0.0523	0.7617	0.9978
Education	2000	48	0.2484	0.0437	0.1530	0.3460
Urbanization	2000	48	0.7141	0.1491	0.3820	0.9440
Age	2000	48	0.5873	0.1492	0.5520	0.6190
Poverty	2000	48	0.0891	0.0282	0.0353	0.1595
PIEF	2000	48	0.0514	0.0415	0.0080	0.2000
PILP	2000	48	0.0283	0.0240	0.0023	0.1105

Table 2 Diversity and economic performance: level vs. growth

Dependent variable	*lnGSP* (2000)			*GSP growth* (1999-2000)	
	(1)	**(2)**	**(3)**	**(4)**	**(5)**
Constant	3.2320****	3.5959****	-0.0047	0.1822**	0.1729**
	(0.3143)	(0.3585)	(0.0990)	(0.1032)	(0.0943)
lnGSP99			0.9986	-0.0381*	-0.0379**
			(0.0260)	(0.0256)	(0.0209)
Race	- 0.0544	0.0113	-0.0271***	-0.0187	-0.0227**
	(0.0676)	(0.0749)	(0.0112)	(0.0145)	(0.0127)
Language	- 0.0662	0.1188	0.0366***	0.0528****	0.0526****
	(0.0919)	(0.0881)	(0.0154)	(0.0146)	(0.0123)
Religion	0.0541	-0.0762	-0.0091	-0.0212	
	(0.1518)	(0.1721)	(0.0251)	(0.0234)	
Education	0.5296***	0.6719****	0.1121****	0.0638*	0.0722**
	(0.2027)	(0.2312)	(0.0351)	(0.0398)	(0.0358)
Urbanization	0.3292****		-0.0049	-0.0033	
	(0.0813)		(0.0159)	(0.0156)	
PopDensity		0.00006*			
		(0.00004)			
Age	1.5151****	1.3282***	0.0081	0.0278	
	(0.5255)	(0.6053)	(0.0951)	(0.0958)	
Poverty				-0.1939****	-0.1813****
				(0.0704)	(0.0649)
*Race *Migration*				0.0002**	0.0002***
				(0.0001)	(0.0001)
*Lang *Migration*				-0.0003**	-0.0003***
				(0.0001)	(0.0001)
N	48	48	48	48	48
Diagnostics					
R^2	0.655	0.5461	0.9908	0.6343	0.6258
Adjusted R^2	0.605	0.4797	0.9893	0.5355	0.5603
Normality	16.4594	14.1327	3.704	0.7598	1.4759
[p-value]	[0.0003]	[0.0009]	[0.1569]	[0.6839]	[0.4781]
Hetero	0.3547	0.4809	0.2017	0.4433	0.4553
[p-value]	[0.9690]	[0.9000]	[0.9985]	[0.9565]	[0.9364]

Notes:
1. Standard errors are in parenthesis and *p*-values for diagnostics tests are in square brackets.
2. ****, ***, **, and * indicates significance level of 1%, 5%, 10% and 20% respectively on the basis of two-tailed tests.
3. *Normality* is the Doornik-Hansen test of normal errors and *Hetero* is White's test for heteroskedasticity

Table 3 Diversity and GSP growth: robustness check (1)

Dependent variable	*GSP growth* (1999-2000)			
	(1)	**(2)**	**(3)**	**(4)**
Constant	-0.1832**	0.1765**	0.1820**	0.1828***
	(0.1003)	(0.0908)	(0.981)	
lnGSP99	-0.0356*	-0.0384**	-0.0339*	-0.0396**
	(0.0250)	(0.0201)	(0.0244)	(0.0198)
Race	-0.0204*	-0.0251***	-0.0177	-0.0243***
	(0.0141)	(0.0122)	(0.0138)	(0.0118)
Language	-0.0168		-0.0209	
	(0.0416)		(0.0360)	
Religion	-0.0222		-0.0205	
	(0.0228)		(0.0223)	
Education	0.0782**	0.0808***	0.0765**	0.0778***
	(0.0396)	(0.0341)	(0.0383)	(0.0335)
Urbanization	-0.0052		-0.0088	
	(0.0152)		(0.0151)	
Age	0.0107		0.0034	
	(0.0937)		(0.0918)	
Poverty	-0.1899****	-0.1800****	-0.1949****	-0.1828***
	(0.0685)	(0.0625)	(0.0670)	(0.0614)
PIEF	0.2191**	0.1686****		
	(0.1233)	(0.0353)		
PLIP			0.4118***	0.2975****
			(0.1854)	(0.0593)
*Race *Migration*	0.0002**	0.0002***	0.0002*	0.0002***
	(0.0001)	(0.0001)	(0.0001)	(0.0001)
*Lang *Migration*	-0.0003*	-0.0003***	-0.0003**	-0.0003***
	(0.0001)	(0.0001)	(0.0001)	(0.0001)
N	48	48	48	48
Diagnostics				
R^2	0.664	0.653	0.678	0.665
Adjusted R^2	0.561	0.592	0.580	0.607
Normality	1.2024	1.659	0.9499	1.2825
[*p*-value]	[0.5482]	[0.4362]	[0.6219]	[0.5266]
Hetero	0.4267	0.6410	0.4700	0.6829
[*p*-value]	[0.9622]	[0.8059]	[0.9428]	[0.7693]

Notes:

1 Standard errors are in parenthesis and *p*-values for diagnostics tests are in square brackets.

2 ****, ***, **, and * indicates significance level of 1%, 5%, 10% and 20% respectively on the basis of two-tailed tests.

3 *Normality* is the Doornik-Hansen test of normal errors and *Hetero* is White's test for heteroskedasticity

Table 4 Diversity and GSP growth: robustness check (2)

Dependent variable	*GSP Growth* (1999-2000)			
	(1)	**(2)**	**(3)**	**(4)**
Constant	0.1589**	0.1089**	0.1354	
	(0.0932)	(0.0636)	(0.0817)	
lnGSP99	-0.0342*	-0.0252**	-0.0289**	-0.0396**
	(0.02067)	(0.0142)	(0.0180)	(0.0221)
Race	-0.02397**	-0.0341****	-0.02676***	-0.0243**
	(0.0124)	(0.0083)	(0.0112)	(0.0131)
Education	0.0727***	0.0885****	0.0702***	0.0778***
	(0.0350)	(0.0238)	(0.0301)	(0.0327)
Poverty	-0.1739****	-0.1017***	-0.1533***	-0.1828***
	(0.0641)	(0.0407)	(0.0603)	(0.0808)
PLIP	0.2899****	0.3373****	0.2772****	0.2975****
	(0.0619)	(0.0356)	(0.0524)	(0.0415)
*Race *Migration*	0.00019**	0.00030****	0.00019***	0.00019****
	(0.00009)	(0.00006)	(0.00008)	(0.00007)
*Lang *Migration*	-0.00028***	-0.00036****	-0.00028***	-0.00028****
	(0.00012)	(0.00006)	(0.00011)	(0.00007)
N	48	48	46	48
Diagnostics				
R^2	0.631	0.428	0.686	0.665
Adjusted R^2	0.566		0.629	

Notes:
1. Standard errors are in parenthesis, *Normality* is the Doornik-Hansen test of normal errors and *Hetero* is White's test for heteroskedasticity
2. ****, ***, **, and * indicates significance level of 1%, 5%, 10% and 20% respectively on the basis of two-tailed tests.
3. Column (2) provides estimates with LAD.
4. The sample in column (3) omits two influential observations with highest absolute value of standardized residuals
5. Only in column (4), heteroskedastic-consistent or robust standard errors are in parenthesis.

Table 5 Endogeniety of *Race*, *Education* and *Poverty*

Instrumented Variable	*Race*		*Education*		*Poverty*	
	(1)	(2)	(3)	(4)	(5)	(6)
Constant	-0.2356*	-1.4199*	-0.2357	-0.0987	0.7229****	0.8152****
	(0.1624)	(1.0235)	(0.4209)	(0.4363)	(0.1960)	(0.1852)
lnGSP99	0.0539*	0.2868	0.1220*	0.0883	-0.1320****	-0.1545****
	(0.0360)	(0.2276)	(0.0921)	(0.0963)	(0.0460)	(0.0436)
Race			0.0465	0.0617	0.0813***	0.0742***
			(0.0630)	(0.0551)	(0.0315)	(0.0320)
Education	0.0781	0.8394***			-0.2729****	-0.2078***
	(0.0613)	(0.3891)			(0.0731)	(0.0822)
Poverty	-00461	1.6005***	0.9469****	-0.8955****		
	(0.1137)	(0.6892)	(0.2536)	(0.2499)		
PLIP	0.4477****	1.9911****	0.3650	0.3355	0.1717	0.1053
	(0.1048)	(0.6135)	(0.2974)	(0.2823)	(0.1609)	(0.1476)
*Race *Migration*	-6.69E-06	0.0027****	-0.0006*	-0.0006	-0.0002	-0.0001
	(0.0002)	(0.0009)	(0.0004)	(0.0004)	(0.0002)	(0.0002)
*Lang *Migration*	0.0003*	-0.0033***	0.0000	0.0001	-0.0000	-0.0002
	(0.0002)	(0.0013)	(0.0005)	(0.0006)	(0.0003)	(0.0003)
Lagged Diversity	1.0137****					
	(0.0223)					
SCI		-0.0816****				-0.00903*
		(0.0224)				(0.00539)
Slavery			0.0295		0.0316*	
			(0.0427)		(0.0225)	
PopDensity00				0.00274		
				(0.00364)		
N	48	48	48	48	48	48
Diagnostics						
R^2	0.994	0.748	0.572	0.574	0.705	0.710
Adjusted R^2	0.992	0.704	0.498	0.499	0.653	0.659
Hausman χ^2	0.13	0.74	6.28	6.06	8.95	6.97
[*p-value*]	[1.000]	[0.9935]	[0.3927]	[0.4164]	[0.1766]	[0.3241]

Notes:

1 Standard errors are in parenthesis and *p*-values for diagnostics tests are in square brackets.

2 ****, ***, **, and * indicates significance level of 1%, 5%, 10% and 20% respectively on the basis of two-tailed tests.

3 *Normality* is the Doornik-Hansen test of normal errors and *Hetero* is White's test for heteroskedasticity

Table 6 Instrumental Variable Estimation

Dependent variable	*GSP Growth*					
	(1)	(2)	(3)	(4)	(5)	(6)
Constant	0.17332**	0.2041*	-0.0151***	-0.0162***	0.2039**	
	(0.0906)	(0.1525)	(0.0068)	(0.0073)	(0.1188)	
lnGSP99	-0.0376**	-0.0436*			-0.0440**	
	(0.0201)	(0.0305)			(0.0248)	
Race	-0.0275***	-0.0217	-0.0453****	-0.0485****	-0.0289**	-0.0418****
	(0.0126)	(0.0206)	(0.0106)	(0.0136)	(0.0172)	(0.0131)
Education	0.0794***	0.0704*	0.1166****	0.1058****	0.0786**	0.0794****
	(0.0338)	(0.0537)	(0.0262)	(0.0299)	(0.0439)	(0.0113)
Poverty	-0.1790****	-0.2095			-0.1944**	-0.0853***
	(0.0619)	(0.1652)			(0.1131)	(0.0352)
PLIP	0.3097****	0.3011****	0.2769****		0.3332****	0.3041****
	(0.0617)	(0.0608)	(0.0631)		(0.0751)	(0.0685)
Language				0.0576****		
				(0.0185)		
Race	0.00021***	0.0002**	0.0002***	0.00028***	0.00031**	0.00028**
**Migration*	(0.00010)	(0.0001)	(0.0001)	(0.00013)	(0.00017)	(0.00015)
Lang	-0.00030***	-0.0003***	-0.00026**	-0.0003**	-0.00042**	-0.00038**
**Migration*	(0.00012)	(0.0001)	(0.00013)	(0.0002)	(0.00022)	(0.00019)
N	48	48	48	48	48	48
Diagnostics						
R^2	0.662	0.664	0.587	0.544	0.652	0.619
Adjusted R^2	0.601	0.605	0.538	0.490	0.591	0.574
Sargan χ^2	3.6703	3.3275	4.519	4.2063	13.0130	13.9583
[*p-value*]	[0.5978]	[0.5046]	[0.9209]	[0.8380]	[0.114]	[0.3768]
R^2 for first stage regressions						
Race	0.995	0.995			0.996	
Poverty		0.733			0.776	
*Race *Migration*					0.660	
*Lang *Migration*					0.676	

Notes:

1. ****, ***, **, and * indicates significance level of 1%, 5%, 10% and 20% respectively on the basis of two-tailed tests.
2. Instruments set used-Column (1)-(4): *lagged diversity*, *SCI*, *Slavery*, *MeanTemp*, *AvegRain* and *AvegPrecip*; Column (5)-(6): *lagged diversity*, SCI, *Slavery*, *MeanTemp*, *AveRain*, *AvePrecip* and the interaction with *MeanTemp*, *AvegRain* and *AvegPrecip* with *Language* and *Race.*

Endnotes

---

[i] Two other related concepts - Social Capability (Temple and Johnson 1998), and Social Infrastructure (Hall and Jones 1999) - complement the rationale for economic outcomes of social capital.

[ii] We use real GSP estimated in chained (2000) dollars that is derived by applying national implicit price deflators to the current dollar GSP estimate for each state.

[iii] For evaluating English fluency of the people who speak a language other than English at home, their self–assessed ability were categorized into four groups: "Very well", " Well", " Not well", and " Not at all".

[iv] The percentage of population reported to choose more than one of the above categories, referred as *Two or more races,* is 2.4 percent (Grieco and Cassidy 2001). The 'race alone' category has been considered to construct the racial diversity index.

[v] According to census definitions, Christians are defined as 'all members, including full members, their children and the estimated number of other regular participants who are not considered as communicant, confirmed or full members' (U.S. Department of Commerce 2003) The Jewish population includes people who define themselves as Jewish by religion as well as in cultural terms. Based on the above, we derive the percent of population who are from other religions.

[vi] In addition to the problem identified in separating race and ethnicity (Sparber 2006) and the subjective evaluation of capturing cultural diversity through ethnic identity as suggested in Ottaviano and Peri (2005), our reservation for measuring ethnic diversity lies with the difficulties in defining different ethnic groups as 22 percent of respondents specified multiple ancestry and 19.1 percent did not report at all (Brittingham and de la Cruz 2004). An index based on number of respondents with single ancestry (58 percent) is most likely to produce a biased estimate and, hence, is not included in this study.

[vii] When the number of groups increases (richness), the diversity index increases and reaches the maximum value of 1 when no pairs of individuals belong to the same group. On the other hand, it reaches the minimum value of 0 when all individuals speak the same language or were born in the same foreign country.

[viii] The maximum value of RQ is 1 with two equally sized groups, and declines as the configuration of groups differs more from this half-half split i.e. the index is decreasing in N. The correlation between religious fractionalisation index and polarization index is 0.724. It is important to note that, preliminary OLS estimates with fractionalisation index does not produce different results and either of polarization or fractionalisation index are not included in the specific model.

[ix] We use the percentage of the population with a bachelor degree or higher as the education variable, rather than a high school degree, as have Shapiro (2006) and Sparber (2006a). Florida (2002) uses the same measure as a talent index and Hendriks (2004) identifies the percentage of population having 16 years of schooling as a measure of 'skill'. Empirical results in the 2004 Hendricks study show that attainment is positively related to measures of agglomeration defined by population size or density.

[x] Grafton et al. (2004) use income inequality or GINI coefficient as a measure for social divergence and provide evidence that the estimated impact of income inequality on total factor productivity level is negative. Based on the studies by Alesina and Ferrera (2002) and Zak and Knack (2001), the negative impact of income inequality can also be transmitted through the reduction in the level of trust in a heterogeneous society.

[xi] After 1980 the foreign-born population has been younger and mostly from Asia and Latin America, while older waves were from mainly Europe. For 1995-2000, the foreign-born population from Africa, had a mobility rate of 68.3 percent, followed by Mexican foreign-born with 62.8 percent, with the least mobile being the European population with 47 percent. The mobility rate for other Latin American population was 57.1 percent and that for Asians was 57.6 percent.

[xii] To evaluate the statistical significance of the interaction effect, we followed the method described in Woodridge (2003). As the mean value of net migration rate of foreign population is 35.235, at the mean *migration* the effect of racial diversity on GSP growth rate is –0.02201 [= -0.02272+0.0002(35.235)] and the effect of linguistic diversity is 0.05154 [=0.0526 + 0.0003(35.235)]. To test the statistical significance of the two partial effects, we rerun the regression after replacing the migration variable with (*migration*-35.235) and derive

the t statistics for the new coefficient. The standard error for the interaction term with racial diversity is 0.00012, which generates t= -0.02201/0.00012=-183.46. The standard error for interaction term with linguistic diversity is 0.00015, which generates t=0.05154/0.00015=343.61

[xiii] Empirical evidence supports the idea that most likely places for knowledge spillovers are essentially the places of professional interactions or informal interactions among colleagues or friends. For skilled individuals, despite having a mother tongue other than English, most of the knowledge exchanges are likely to take place in English.

[xiv] About half of the sample population i.e. 42 percent reported a category other than single ethnicity and 19.9 percent remain unclassified or not reported at all.

[xv] In the Simon and Nardinelli (2002) study, the coefficient for manufacturing variable appears as being statistically significant and positive in 1900-20 and 1920-40 growth regressions, but negative for the period 1960-86.

[xvi] Ottoviano and Peri (2005) argue that as Spanish and Chinese cultures are more different than Spanish and Italian cultures. We conclude that communication in case of the former pair will have higher communication costs, in comparison with than that of the second pair.

[xvii] According to the US Census glossary, a household in which no person 14 years old or over speaks only English and no person 14 years old and over who speaks a language other than English speaks English “very well” is classified as “linguistically isolated”. In other words, a household in which all members 14 years old and over speak a non-English language and also speak English less than “very well” is “linguistically isolated”.

[xviii]Under robust regression, lower weights are assigned to Oregon and Kentucky of 0.4353 and 0.5497 respectively. Utah, for example, receives a weight of 0.8755. Oregon and Kentucky are also interpreted as potential outliers for specific models in column (5), Table 2 and column (2), Table 3.

[xix] The major categories for measuring SCI are (Putnam 2000:291): i measures of community organizational life; ii measures of engagement in public affairs; iii measures of community volunteerism; iv measures of informal sociability; and v measures of social trust.

[xx] Olson-type organizations are political organizations, professional organizations, business organizations, and labour organizations. Putnam-type organizations are bowling centres, civic and social associations, physical fitness facilities, public golf courses, religious organizations, sports clubs, managers and promoters.

[xxi] The website address for this dataset is: http://fisher.lib.virginia.edu/collections/stats/histcensus.

[xxii] “Slavery was, in fact, a social system designed to destroy social capital among slaves and between slaves and freemen. Well-established networks of reciprocity among the oppressed would have raised the risk of rebellion, and egalitarian bonds of sympathy between slave and free would have undermined the very legitimacy of the system”, so writes Putnam (2000:294).

[xxiii] The data is available online: http://www.arts.york.ca/econ/lagerloef/PubDataUSCan.dta.

[xxiv] The Hausman test statistics, χ^2 (4) = 17.11 and Prob>χ^2=0.0018, when *Race, Language* and *Poverty,* all are considered endogenous.

[xxv] The Hausman test statistics, χ^2 (5) = 3.61 and Prob>χ^2=0.6072, when only *Race* and *Poverty* are considered endogenous.

[xxvi] See Hum and Simpson (2001), Meng and Gegory (2005), and Stolzenberg and Tienda (1997) for different versions of definitions of economic assimilation.