Does Multiculturalism Pay? Empirical Evidence from the United States and Canada*

Nazmun N. Ratna,¹ Quentin Grafton² and Ian A. MacDonald¹

We investigate the economic impacts of social diversity and the consequent barriers of communication in Canada and the United States. Social diversity is explained by linguistic, cultural and religious differences across the 48 contiguous states in the United States and the 10 provinces in Canada. The ordinary least squares (OLS) and instrumental variables estimation show that social diversity increases per capita gross domestic product at the state and province level, but the positive economic pay-off from diversity diminishes as the level of fluency in official language declines. The empirical results provide an important economic rationale for overcoming linguistic divisions and “inclusive” multiculturalism in other pluralistic countries, such as Australia.

Keywords: cultural diversity, linguistic diversity, multiculturalism, social inclusion.

1. Introduction

This study evaluates the economic impact of social diversity in Canada and the United States. It provides an empirical analysis as to the economic impact of diversity. Some research, supported by a significant number of cross-country studies, suggests that ethno-linguistic diversity is detrimental to economic growth and the optimal allocation of resources. By contrast, others conclude that in urban centres with high human capital, diversity contributes to economic growth via knowledge diffusion and innovation. We empirically investigate the basis for causality between economic pay-off and social diversity. In particular, we test the hypothesis, first postulated by Grafton *et al.* (2004, 2007), that the economic pay-off to social diversity, as measured by regional per capita income, is diminished by barriers to communication across different social groups.

We measure social diversity in terms of language, religion and culture and obtain two robust and statistically significant results: first, diversity has a positive impact on per capita Gross State Product (GSP) in the United States and per capita Gross Provincial Product (GPP) in Canada; and second, the beneficial effects of diversity decline in the presence of linguistic isolation. These findings are robust to alternative model specifications even after accounting for the possibility of reverse causality between social diversity and state and provincial per capita income.

The remainder of the article is structured as follows: Section 2 provides motivation for this study and a brief literature review; Section 3 describes the data and variables for the econometric model;

*We thank the two anonymous referees and conference participants at the Western Economic Association International (WEAI) Conference in Vancouver for their helpful comments. We also thank Huy Hoang and Shuzhang Sun for their assistance in comprising the datasets, and gratefully acknowledge Lincoln University for providing funding to make this possible.

¹Department of Accounting, Economics & Finance, Faculty of Commerce, Lincoln University, New Zealand.

²Crawford School of Economics and Government, Australian National University, Australia.

JEL classifications: C21, O51, R11

*Correspondence: R. Quentin Grafton, Crawford School of Economics and Government, ANU College of Asia and the Pacific, Australian National University, Canberra, ACT 0200, Australia. Email: quentin.grafton@anu.edu.au

Section 4 outlines the econometric model and presents the initial estimation results; Section 5 reviews a series of robustness checks of the model; Section 6 details model estimates that account for endogeneity of key variables; Section 7 discusses the policy implications of the findings; and Section 8 concludes.

2. Motivation

Easterly and Levine (1997) identify diversity, measured by ethno-linguistic fractionalisation, as a causal factor for differences in economic performance across countries. They argue that higher diversity is detrimental to economic growth because of a conflict of preferences over the allocation of resources. This work has motivated a series of studies focussed on reduced interpersonal trust (Alesina and Ferrara, 2002; Leigh, 2006a,b), conflict of preferences over redistributive policies (Alesina *et al.*, 2003; Alesina and Ferrara, 2005), reduced allocation for “productive public goods” (Alesina *et al.*, 1999) and lack of political rights (Collier, 2000). Grafton *et al.* (2004, 2007) highlight the economic significance of social barriers to communication. In their study, they show that higher linguistic fractionalisation reduces Total Factor Productivity (TFP) growth due to the lack of intergroup communication and contend that such “bridging” communication is essential to maximise the potential for knowledge spill over and innovation.

Bellini *et al.* (2008) provide a useful summary of the literature on economic impacts of social diversity. In the early literature, Florida (2002) and Grafton *et al.* (2002), among others, provide evidence that social diversity contributes to productivity and employment growth. The central theme of this research is that diversity may increase the potential for innovation and creativity by mixing abilities, experiences and cultures in a heterogeneous society. In addition, Grafton *et al.* (2002) suggest that comprehensive schooling, cross-cultural exchanges and support for common national languages can reduce social barriers to communication that promote productivity-enhancing knowledge exchange and innovation. In a study on labour productivity across US cities, Ottaviano and Peri (2005) provide evidence that the impact of diversity on wages of native workers is positive. Bellini *et al.* (2008) generate similar results using data from 12 countries in the EU region.¹

The finding that diversity can have positive economic pay-offs contrasts to the results of Borjas (1995, 2003) who found that the impact of immigration on wages is negative. Furthermore, Peri and Sparber (2009) compared the task intensity of occupations for the US labour market for 1960–2000. They conclude that foreign and native workers are not perfect substitutes because the two groups tend to specialise in different kinds of work, that is foreign workers specialise in manual tasks and native workers on interactive tasks. As a result, a large influx of immigrant workers competing for manual jobs will be detrimental for low-educated native workers.

In this study, we empirically estimate the economic significance of social barriers to communication. Our contribution is threefold. First, we use a previously unused measure for linguistic barriers that quantifies low levels of fluency in the official language(s) and test for its statistical and economic significance in explaining levels of economic performance. Such an analysis is important from a policy perspective because, in recent years, there has been an increasing preference in high-income English-speaking countries, including Australia, to attract immigrants with a high level of English language fluency.

Second, our investigation provides evidence-based insights for policy formulation in both Canada and the United States. This is important because in 1971 Canada adopted an official policy of multiculturalism and, in 1988, became the first country in the world to pass a multiculturalism law (Kymlicka, 2003a,b). Although the United States does not have an official policy on “multiculturalism”, it is, nevertheless, culturally pluralistic. Thus, empirical insights based on data from both countries may provide opportunities for state/provincial and national governments in Canada and the United States to increase the benefits that may accrue from a multicultural/pluralistic society.

¹The countries are Austria, Belgium, Denmark, France, former Western Germany, Ireland, Italy, the Netherlands, Portugal, Spain, Sweden and United Kingdom.

Third, we are the first to combine US and Canadian data in the linguistic barriers context and to compare results between the two countries. This provides a “natural experiment” in the policy setting because both share a similar history of early European settlement, and the political institutions that accompanied this early settlement. Thus, our work helps to overcome a major limitation of many previous cross-country studies where “like” is not compared with “like.” Although the two countries’ do differ, especially in terms of population size, their demographic profiles are similar, both possess broadly similar cultural origins of pre-colonial and colonial populations, and each has received large subsequent influxes of migrants. Consequently, we are able to compare possible economic differences associated with multiculturalism (Canada) versus cultural pluralism (United States) that may relate to linguistic barriers.

3. Data and Variables

3.1 Regional Income and Diversity Variables

We use a cross-sectional dataset comprising the 48 contiguous states in the United States and the 10 Canadian provinces. The US data is for 2000 and the Canadian data is for 2001.² To construct the dependent variable, we combine GSP from the Bureau of Economic Analysis, US department of Commerce, and GPP from Statistics Canada. Per capita real GSP is defined as the sum of the value added originating in all industries within the state, divided by its population. We derive a purchasing power parity-adjusted value for per capita real Gross Province Product (GPP) as a counterpart of GSP in the Canadian context.

We derive fractionalisation measures by grouping people according to language, region of birth and religion. Given the interwoven layers of social identity, and consequent varied patterns of interaction, finding a “right” variable or group of variables for measuring diversity and fractionalisation is a challenge. For example, someone born in Canada to a Lebanese father and Italian mother may speak French and also be a Buddhist. Or despite the difference in race, a second-generation Scandinavian American and second-generation African American born in the United States may both identify themselves as American. Moreover, interactions between diverse groups accentuate the problems by further blurring the lines of distinction. Nonetheless, we contend that the likelihood of knowledge interaction is greater when sociocultural proximity is closer due to people sharing similar language, religion and region of birth.

Although social divisions may still be evident across race and ethnicity, especially in the United States, these are not included in our data set.³ Using data on race and ethnicity would be problematic for two reasons. First, race is typically used to represent the geographical area from where an individual’s outward physical attributes are most evident, and ethnicity tends to be defined more by the cultural differences that are shaped by national identity, culture and mother tongue. But problems arise when, for example, according to the US census, both Europeans and Arabs are defined under the same category for race (White) despite differences in ethnicity, language and culture between the groups. Or someone from Madras, India, could be “lumped” together with someone from West Bengal under the same category of ethnicity, despite their differences in culture and mother tongue. Second, reliance on self-reporting would further complicate matters if groups differ in their perceptions of race and ethnicity. For example, the US census and the US National Research Council both consider Hispanics as a separate ethnic group, but Hispanics typically identify themselves as a separate race.

Language is arguably the single most important determinant of social identity in any society. In this article, linguistic diversity, that is *Language*, for the United States is based on US Bureau of Census information on the responses for *language spoken at home*. Respondents (5 years and over) were asked if they “speak a language other than English at home,” and their responses were compiled under five broad categories: English, Spanish, other Indo-European languages, Asian and Pacific Island Languages, and

²Full set of data for 2010 US Census is yet to be published.

³Province-level data on race for 2001 is not available for Canada.

all other languages. For the Canadian provinces, we derive linguistic diversity based on Statistics Canada data on Mother tongue. For Canada, we identify French as a separate linguistic category. To address the significance of official bilingualism in Canada, we identify three major linguistic groups; Anglophone, Francophone and others.⁴ We elaborate on bilingualism in Canada in Section 6.

We measure religious diversity for two reasons. First, churches or mosques provide the basis for “bonding” of social capital for groups of people who share religious beliefs, but are otherwise diverse. Second, in recent times, social cleavages are drawn along religious divides, especially if there are numerous immigrants with diverse religious backgrounds. US Bureau of Census contains information on only two major categories of religion for 2000, Christian and Jewish. To allow for comparability between the US and Canadian data sets, religious diversity, that is *Religion*, is compiled into three groups for both countries: Christian, Jewish and others.⁵

A third measure of social diversity, *Culture*, is used to capture the likelihood of group formation based on cultural proximity. *Culture* is constructed on the basis of a person’s region of birth: Europe, Asia, Africa, Oceania, Latin America and North America. This diversity measure is likely to capture other residual dimensions of social diversity and helps to verify the robustness of the estimates for the other two diversity measures (language and religion).

We derive the diversity measure as a fractionalisation index that is commonly applied in the social diversity and growth literature. The fractionalisation index given in Equation 1 is a *Herfindahl* concentration index and signifies the likelihood that two people chosen at random from a diverse universe will belong to different groups. Diversity is defined as:

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

In Equation 1 f_{ji} is the share of group (linguistic, religious, racial, or birth region of foreign population) j ($j = 1, 2, \dots, n$) in state i . The minimum value for this index (0.00) indicates complete homogeneity and its maximum value (1.00) represents complete heterogeneity, that is every individual belongs to a different group. According to our dataset, Newfoundland and Labrador have the lowest linguistic diversity of 0.0316, whereas California has the highest linguistic diversity of 0.5578.

3.2 Control Variables

We select a set of control regressors in our model that are supported by empirical evidence from other studies on regional differences across the United States. All of these controls are primarily derived from the US Bureau of Census and Statistics Canada. We include a control variable, *Education*, to denote differences in terms of educational attainment across States or provinces. A significant number of studies of the US economy show that differences in human capital are associated with differences in per capita income or productivity across states or counties (Beeson, 1987; Ciccone and Hall, 1996; Hendricks, 2004; Sedgley and Elmslie, 2004; Higgins *et al.*, 2006), and cities or metropolitan areas (Glaeser *et al.*, 1992; Rauch, 1993; Ciccone and Hall, 1996; Rigby and Essletzbichler, 2002; Simon and Nardinelli, 2002; Shapiro, 2006). We define *Education* as the percent of the population with a Bachelor’s degree or higher.⁶

We address the causal link between economic performance and agglomeration variables like population density or urbanisation, as was first envisaged by Alfred Marshall. Marshall emphasises labour market externalities generated from urbanisation due to skill concentration. Subsequent literature focuses on the role of human capital externalities in the urban context, especially in terms of employment and innovation (Glaeser *et al.*, 1992; Jacobs, 1969 cited in Fu, 2007; Ottaviano and Peri, 2005). Closely linked with Marshall’s labour market externalities is the contribution of creative capital for urban growth. Florida (2002) identifies the “creative class,” as a combination of two groups of people:

⁴We examine with different definition of linguistic groups in Section 4.

⁵The OLS estimates with eight religious groups, as specified by Statistics Canada, are qualitatively same.

⁶Florida (2002) uses the same measure as a talent index.

one, “professionals such as doctors, lawyers, scientists, engineers, university professors; and two, bohemians,” made up of artists, musicians and sculptors’. Florida (2002) argues that both groups contribute to new innovation and technological development. We argue that urbanisation, or high population density, enables frequent interactions between people that enhances knowledge transfers and innovation. Hence, we include *Urbanisation*, defined as the percentage of population in metropolitan areas as a proxy for agglomeration.⁷

Cross-country studies indicate that political variables such as strength of institutions (Acemoglu *et al.*, 2001, 2002; Rodrik *et al.*, 2004) and degree of democratisation (Collier, 2000) influence the level of economic performances across countries. As the states or provinces are similar in terms of structure and strength/quality of institutions, we consider two indexes of economic freedom developed by the Fraser Institute as proxies for differences in the policy environment. An economic freedom index captures the impact on restrictions in economic freedom. Two indices are constructed on a 10-point scale: (i) a subnational index (based on state/provincial and local/municipal governments), that is Economic Freedom at State-level or *EFS*; and (ii) a government index (inclusive of federal, state/provincial and local/municipal levels); that is Economic Freedom at Federal-level or *EFF*. They are derived from the relative ranking on 10 variables under three major categories: size of the government; takings and discriminatory taxation; and labour market reform.

We extracted data from the sixth edition of *The Economic Freedom in North America*. The report suggests that econometric estimations show that a 1 per cent increase in *EFF* will result in an increase of 1.02 per cent in the growth rate of GSP per capita, and an increase of 0.60 per cent in the growth rate of GPP per capita. The same increase in *EFS* will result in an increase of 0.73 per cent in the growth rate of GSP per capita, and an increase of 0.56 per cent in the growth rate of GPP per capita. Thus, we include both *EFF* and *EFS* as control variables.

Appendix III presents a correlation matrix for endogenous variables of the model. *Language* has strong correlation with urbanisation, but is not highly correlated with *Education* or any of the economic freedom variables, that is *EFF* and *EFS*. As expected, *EFF* and *EFS* are highly correlated with each other and each has strong correlation with *Education*. *Religion* is strongly correlated with *Education*, and hence with *EFF* and *EFS*. *Culture* does not show a strong correlation with any of the variables. Although there are debates about the “right” level of correlation to identify the independent effect of each of the endogenous variables, none of the correlation coefficients with the variables of interest, that is fractionalisation indexes, indicates multicollinearity to be a potential “problem.”

4. The Empirical Model

The empirical model used to test the hypothesis that barriers to communication reduce the positive impact of diversity is provided in Equation 2. We estimate several variants of Equation 2, in which the dependent variable is the natural logarithm of per capita GSP or GPP for the *i*-th state/province for a particular year *T*:

$$\ln Y_{iT} = \pi + \alpha \text{Language}_i + \beta \text{Region}_i + \delta \text{Culture}_i + \gamma \text{LB}_i * \text{Language}_i + \theta \text{Control}_i + \varepsilon_i Y_{it} \quad (2)$$

where *Language*, *Religion* and *Culture* are the three fractionalisation measures discussed earlier and *Control* is a vector of other explanatory variables influencing per capita GSP/GPP. To capture the importance of communication, both between immigrants and natives and between different immigrant groups, we include an interaction variable between *Language* and *Linguistic Barrier (LB)* where the latter variable is a proxy for barriers to communication and is measured as the level of proficiency in the official language.

Data on fluency in official language differ between the United States and Canada. Hence, we use two interpretations of linguistic barriers. The US census measures English fluency for people who do not speak English at home. This is a self-assessed ability categorised into four groups: “Very well,”

⁷In our initial estimation, we tested with another agglomeration variable, population density. The coefficient for population density was not statistically significant, and the OLS estimates for other variables were similar to column 1, Table 1.

Table 1. Diversity and Economic Performance: OLS

	ln (GSP/GPP per capita)				
	(1)	(2)	(3)	(4)	(5)
Constant	8.8356*** (0.2195)	8.8602*** (0.2027)	8.2873*** (0.1771)	8.2237*** (0.2326)	8.3811*** (0.2367)
<i>Language</i>	1.5391*** (0.3222)	1.0762*** (0.3418)	1.5750*** (0.1871)	1.7903*** (0.2172)	1.1360*** (0.3570)
<i>Religion</i>	0.8633*** (0.2681)	0.7229*** (0.2558)	0.5436*** (0.1994)	0.6441*** (0.2352)	0.4796** (0.2399)
<i>Culture</i>	0.9979*** (0.2880)	0.9847*** (0.2376)	1.1178*** (0.1768)	1.3555*** (0.2110)	1.0473*** (0.2600)
<i>Language * LB</i>	-7.0980*** (1.7613)	-5.8204*** (1.6759)	-6.9985*** (1.1693)	-7.4532*** (1.3736)	-5.3552*** (1.6361)
<i>Canada_Dummy</i>	0.4219*** (0.1235)	0.2899** (0.1209)	0.4698*** (0.0905)	0.4511*** 0.1066	0.3856*** (0.1110)
<i>Education</i>	0.6786 (0.5313)				0.4637 (0.4494)
<i>Urbanisation</i>		0.4959*** (0.1702)			0.3290** (0.1586)
<i>EFF</i>			0.1255*** (0.0184)		
<i>EFS</i>				0.0918*** (0.0211)	0.0768*** (0.0214)
N	58	58	58	58	58
<i>Diagnostics</i>					
<i>R</i> ²	0.6245	0.6678	0.7970	0.7176	0.7453
Adjusted <i>R</i> ²	0.5803	0.6287	0.7731	0.6844	0.7037

Notes: Standard errors are in parenthesis. ***, ** and * indicates significance level of 1 per cent, 5 per cent and 10 per cent respectively on the basis of two-tailed tests.

Source: Authors' compilation with data from US Bureau of Census and Statistics Canada

“Well,” “Not well” and “Not at all.” A measure for linguistic isolation is then derived from these data. For the United States, *LB* is defined as the percentage of linguistically isolated household per states.⁸ For Canada, we define *LB* as the percentage of “*allophone*” workers who use a language other than English or French most often or regularly at work. *Allophone* workers are defined as individuals aged 15 years and over, whose mother tongue is a language other than English or French (*Use of English and French at work*, 2001 Census: Analysis series, Statistics Canada).

To test the validity of our control variables, we first estimate the model with each control variable included on its own. Column (1), Table 1 provides OLS estimates for Equation 2 when only *Education* is included as a control variable. The estimated coefficients for all of the causal variables, *Language*, *Religion* and *Culture*, have a positive sign and are statistically significant at the 1 per cent level. The positive sign for the diversity measures support the hypothesis that diversity contributes positively towards GSP/GPP per capita. The estimated coefficient of the interaction variable, that is *Language*LB*, has the hypothesised negative sign and is statistically significant at 1 per cent level. This result provides support for the hypothesis that although linguistic diversity increases per capita GSP/GPP, its effect is decreased because of linguistic barriers. The coefficient for *Education* has the hypothesised sign, but is not significant at the conventional statistical levels. The estimated coefficient has a *t*-statistic of 1.28 with a *P*-value of 0.207. We also include a dummy variable for Canadian provinces, that is *Canada_dummy*. The coefficient for *Canada_dummy* is significant at 1 per cent level. This suggests that differences in the two countries may have an impact of state/provincial-level economic performance. We discuss the policy differences in Section 7.

Columns (2), (3) and (4) present the results for when *Urbanisation* (Column 2), *EFF* (Column 3) and *EFS* (Column 4) are included as the sole control variable. In each case, the estimated coefficients for the controls are positive, as hypothesised, and are statistically significant at the 1 per cent level. The estimated coefficients for all other variables are largely unchanged compared to the results in Column (1).

Column (5), Table 1 reports the estimated coefficients for Equation 2 with the full set of control variables. The coefficients for *Language* and *Culture* retain their positive signs and are statistically significant at the 1 per cent level. The coefficient for *Religion* retains its positive sign, but is statistically significant only at the 5 per cent level. The interaction variable, *Language*LB* retains its hypothesised negative sign and is statistically significant at the 1 per cent level.

Tests were undertaken for heteroscedasticity and potential for omitted variables. The Breusch–Pagan test for heteroscedasticity reports a χ^2 value of 2.06 with a *P*-value of 0.1514. Thus, we cannot reject the null hypothesis of homoscedasticity. We also conducted White’s test for heteroscedasticity, which validates homoscedasticity with a χ^2 test statistic (46.08) with a *P*-value of 0.3461. We derive a Ramsey RESET test statistic to verify the potential for omitted variables. The F-statistic is 0.54 with a *P*-value of 0.6575. Thus, we cannot reject the null hypothesis that the model (Column 5, Table 1) has no omitted variables. We check the robustness of these results in the following section.

5. Robustness of OLS Results

In this section, we check the robustness of the OLS results. We first use an extended set of control variables and then apply a set of alternative specifications to check the influence of potential outliers. Finally, we adopt an alternative econometric estimation, that is general to specific modelling, using the PcGets algorithm.

5.1 Robustness Check: Potential Omitted Variables

This section looks at the potential problem of omitted variables that may influence the regional income differences. To verify the robustness of the basic OLS model (Column 5, Table 1), we include

⁸According to US census, 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 defined as a “linguistically isolated” household. 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.”

additional control variables and estimate the model. If a particular additional control variable is statistically significant, it is retained for the next set of estimations, otherwise it is removed.

First, we introduce dummy variables to capture the differences in climate, natural resources, infrastructure and other historical factors across the regions. We follow the US Bureau of Census definitions by recognising four regions: Northeast, Midwest, South and West. Appendix IV provides the list of states and provinces for each region. None of these regional variables were statistically significant when added to the basic model in Column (5), Table 1.

Second, we define New York, California, Florida, British Columbia and Ontario as “anchor” states/provinces, and introduce a dummy variable named *Anchor-dummy*. All of the anchor states have a higher concentration of immigrants, and with the exception of Ontario, all are coastal states. Coastal states tend to be the first port of entry for immigrants and tend to be the most economically advanced states/provinces. The estimated coefficient for *Anchor-dummy* is not statistically significant.

Third, we consider a dummy variable that accounts for regional differences in the Hispanic population across US states. We name this variable *Latin-dummy* to denote the states, including Arizona, California, Florida, New Mexico and Texas, that are in close proximity to Mexico and the other Spanish-speaking countries of Latin America. The coefficient for *Latin-dummy* is not statistically significant from zero.

We use an alternative measure for linguistic diversity, based on a different set of language groups, to further evaluate the importance of this variable. Ottaviano and Peri (2005) argue that languages with a common linguistic root indicate similarities in culture. To include this aspect of cultural proximity due to common linguistic roots, *Language 1* is derived as a fractionalisation index based on eight major linguistic groups: i. Anglo-Saxon or Germanic; ii. Neo-Latin; iii. Slavic; iv. Indo-Iranian; v. Greek; vi. Asian; vii. Native American; and viii. Other languages. Column (1), Table 2 provides the estimates after the inclusion of *Language 1* as a control variable.⁹ The estimated coefficient for *Language 1* is significant at the 1 per cent level and the coefficient on the interaction variable with *LB* is significant at the 1 per cent level with the hypothesised negative sign. The coefficients for *Culture* and *Religion* are significant at the 1 per cent level and 5 per cent level respectively. The estimated coefficients for the other variables are qualitatively similar to previous estimates in Column (5), Table 1.

5.2 Robustness Check: Biased Estimation

Column (1), Table 2 provides a diagnostics for heteroscedasticity, denoted as *Hetero*. This is a test statistic for White’s test, a χ^2 distribution, and reported as 0.4608 with a *P*-value of 0.9440. Thus, we cannot reject the null hypothesis of homoscedasticity for the model specification with *Language 1*. Column (1), Table 2 also reports the Doornik–Hansen Test, with a *P*-value of 0.0164. We conduct the same test for the final basic model (Column 5, Table 1). The Doornik–Hansen test statistic for final basic model is 8.1808, with a *P*-value of 0.0167. For both specifications, we reject the null hypothesis of normality of errors and evaluate the influence of potential outliers in this section.

Column (2), Table 2 reports estimated coefficients for *Robust Regression* Analysis. None of the observations in our dataset is identified as a gross outlier by Cook’s distance measure, so the robust regression does not exclude any observations. Using robust regression, we find that there is a very small reduction in value in the estimated coefficients for *Culture* and the interaction term (*Language * LB*) and a slight increase in the estimated coefficients for *Language* and *Religion* in comparison to Column 5, Table 1. The coefficients for the three diversity measures retain their positive sign. The estimated coefficients for *Language* and *Culture* are statistically significant at the 1 per cent level and the coefficient for *Religion* is statistically significant at the 5 per cent level. The interaction variable retains its negative sign and is statistically significant at the 1 per cent level (Column 2, Table 2). Column (3), Table 2 presents the results of a quantile or median regression. As with the previous estimates (OLS and robust regression estimation), the results with quantile regression are qualitatively similar.

⁹The correlation coefficient between *Language* and *Language 1* is 0.9938.

Table 2. Diversity and Economic Performance: Robustness Checks

	(1)	(2)	(3)	(4)	(5)
	ln(GSP/GPP per capita)				
Constant	8.4077*** (0.1004)	8.4603*** (0.2261)	8.3906*** (0.2635)	8.3272*** (0.2311)	8.3570*** (0.2274)
<i>Language</i>		1.3384*** (0.3055)	1.5894*** (0.3078)	1.3150*** (0.3122)	
<i>Religion</i>	0.4753** (0.2430)	0.5611** (0.2292)	0.7316*** (0.2341)	0.5339*** (0.2342)	0.5279** (0.2391)
<i>Culture</i>	1.0410*** (0.2601)	1.1464*** (0.2148)	1.1440*** (0.2417)	1.1913*** (0.2195)	1.1954*** (0.2215)
<i>Language*LB</i>		-5.8816*** (1.4671)	-6.7687*** (1.4740)	-6.0336*** (1.4991)	
<i>Urbanisation</i>	0.3455** (0.1574)	0.2836* (0.1553)	0.2270 (0.1689)	0.3273** (0.1587)	0.3475*** (0.1578)
<i>EFS</i>	0.0746*** (0.0213)	0.0653*** (0.0209)	0.0629** (0.0244)	0.0792*** (0.0213)	0.0770*** (0.0212)
<i>Canada_dummy</i>	0.3647*** (0.1098)	0.3440*** (0.1079)	0.4032*** (0.1185)	0.3719*** (0.1103)	0.3427*** (0.1083)
<i>Language 1</i>	1.0688*** (0.3406)				1.2485*** (0.3016)
<i>Language 1*LB</i>	-5.1038*** (1.5332)				-5.7729*** (1.4168)
N	58	58	58	58	58
<i>Diagnostics</i>					
<i>R</i> ²	0.7462		0.5179	0.7398	0.7396
Adjusted <i>R</i> ²	0.7047			0.7034	0.7032
Normality	8.2168			6.1999	5.8515
[<i>P</i> -value]	[0.0164]			[0.0451]	[0.0536]
<i>Hetero</i>	0.4608			0.5299	0.4841
[<i>P</i> -value]	[0.9440]			[0.8903]	[0.9194]

Notes: Standard errors are in parenthesis. Columns 1, 2 and 3 report OLS estimation, robust regression and median regression respectively. Column 4 and 5 report specific models for Column 5, Table 1 and Column 1, Table 2 respectively. *Normality* is the Doornik-Hansen test of normal errors and *Hetero* is White's test for heteroscedasticity. *R*² in Column 3 represents *Pseudo R*². ***, ** and * indicate significance level of 1 per cent, 5 per cent and 10 per cent respectively on the basis of two-tailed tests.

Source: Authors' compilation with data from US Bureau of Census and Statistics Canada.

5.3 General to Specific Modelling

To further check the robustness of our OLS results, we employ a General to Specific Modelling or *Gets* modelling that extends the “LSE approach” (Krolzig and Hendry, 2001). *Gets* modelling uses the *PcGets* algorithm to derive a “specific model” with all statistically significant variables. It starts with a “General Unrestricted Model (GUM),” which is typically a standard OLS specification and then follows a multi-step procedure to select the “specific” model. At each step, *PcGets* uses a set of diagnostic tests for two criteria: congruency and encompassing. The *PcGets* algorithm provides a robustness check in small datasets to identify whether the hypothesised variables of interest are selected in a final model that is determined by the statistical properties of the variables rather than the model choices of the researchers.

We chose the models represented in Column (5), Table 1 and Column (1), Table 2 as the *general* unrestricted models. The results in Columns (4) and (5), Table 2 are the estimates of the *specific* model based on the *PcGets* algorithm. The coefficients for *Language* and *Culture* are significant at the 1 per cent level, and the coefficient for *Religion* is significant at the 5 per cent level. The coefficients for *Urbanisation* and *EFS* retain their hypothesised signs in the *specific* model and are statistically significant at the 1 per cent level.

For the specific models in Columns (4) and (5), Table 2, the statistical significance of the estimated coefficients of two variables is noteworthy. First, the estimated coefficient for the *Canada dummy* is statistically significant at the 1 per cent level of significance. The estimated coefficient for *Canada dummy* in Column (4) implies that Canadian provinces have 37.19 per cent higher per capita GPP than US states with the same level of social diversities, English inefficiency, urbanisation and economic freedom. Second, the interaction terms with linguistic diversity (i.e. *Language* or *Language 1*) and barriers to communication (i.e. *LB*) are statistically significant at the 1 per cent level in the specific models.

To evaluate the partial effect of linguistic diversity, we evaluate the interaction term at the mean value of *LB*, that is 0.06695, and also test the statistical significance of the partial effect. First, at the mean value of *LB*, the partial effect of linguistic diversity is 0.9111 [=1.3150–6.0366(0.06695)]. Column (5), Table 1 and Column (4), Table 2 report that the coefficients for linguistic diversity, 1.1360 and 1.3150 respectively, and are higher than the estimated partial effect of linguistic diversity in the presence of barriers to communication. Second, we rerun the regression replacing *LB* with *LB*–0.06695. Based on the t-statistics for the new coefficients, *Language* is statistically significant at the 1 per cent level and its partial effect on per capita GSP/GPP is 0.9111.

We also tested for heteroscedasticity. Table 2 reports the test statistic for White’s test, a χ^2 distribution defined as *Hetero*. As with the general models with OLS estimation, *Hetero* test statistics for specific models are statistically insignificant with a *P*-value of 0.8903 (Column 4, Table 2) and 0.9194 (Column 5, Table 2). Thus, we cannot reject the null hypothesis of homoscedasticity for the specific models.

To test for possible biases that may arise from omitted variables for the *specific* model in Column (4), Table 2, we derive Regression Specification Error Test (RESET) test-statistics. The *F*-statistics is 0.55, with a *P*-value of 0.6517. Hence, the null hypothesis cannot be rejected and we conclude that there is no evidence for omitted variables.

In summary, we find that the positive impact of social diversity on per capita GSP/GPP is reduced with the presence of barriers to communication, defined as a low level of fluency in official languages. We argue that this lack of fluency increases the cost of communication across the subgroups of population, and consequently affects knowledge interactions and the level of economic performance.

6. Instrumental Variable Estimation (IVE)

A possible problem with our OLS estimates is that the three social diversity measures are endogenous. In other words, states or provinces with higher per capita incomes such as California or Ontario, or cities such as New York or Toronto, are more likely to attract non-native workers because of better economic opportunities. To address this issue and derive the rationale for IVE, we conducted three

statistical tests: a Hausman test to verify if the OLS results are consistent; a reduced-form equation to check the excludability/validity of instruments; and a *Sargan* test to ascertain if instrumental variables are correlated with the error term.

First, we use *Slavery*, that is, percentage of slave population in 1850, as an instrument. As social segregation was institutionalised under slavery, racial segregation in the 18th century has largely shaped social cleavages in the United States (Putnam, 2000). Although racial diversity is *not* included in our analysis because of data unavailability, we argue that social divisions across groups may still be influenced by racial separation. The *Hausman* test statistic for *Slavery* is 13.41 with a *P*-value of 0.0369. Thus, we reject the null hypothesis that OLS results are consistent. In the reduced-form equation for *Language*, the estimated coefficient for *Slavery* has the hypothesised negative sign and has a *t*-statistic of -1.44 (*P*-value of 0.156). However, the coefficient for predicted error in the structural equation is not statistically significant and has a *t*-statistic of -1.25 , with a *P*-value of 0.217. Hence, *Slavery* is considered to be a valid instrument for linguistic diversity.

We checked for endogeneity of *Language* using different instruments, that is a set of geography variables: average/mean temperature (*AvgTemp*) and its square (*AvgTempSq*), average precipitation (*AvgPrecip*) and average number of rainy days per year (*AvgRain*).¹⁰ These variables are chosen on two grounds. First, Sachs (in Rodrik *et al.*, 2004) provides empirical evidence that variables such as climate, endowment of natural resources and disease influence productivity growth. Second, geography contributed to the choice of early settlement by Europeans. The East Coast of North America hosted the huge influx of early settlers from Europe in the 19th century. Schools and businesses later expanded around the densely populated area that contributed to human capital accumulation and, thereby, economic growth. As a result, geographical factors may influence current social diversity, but are not correlated with GSP/GPP per capita at the beginning of the 21st century.

We grouped the geography variables into two parts; one, *AvgTemp* and *AvgTempSq* and two, *AvgPrecip* and *AvgRain*. For *Language*, the Hausman test statistics provide mixed results with 9.75 (*P*-value of 0.135) for group 1 and 18.23 (*P*-value of 0.000) for group 2 variables. As we reject the null hypothesis at the 1 per cent significance level for group 2 variables, we estimated the reduced-form equation with these variables only. The coefficient of predicted error in the structural equation has *t*-statistic of -0.38 , and a *P*-value of 0.708. This implies that *AvgPrecip* and *AvgRain* do not appear to have an independent effect on GSP/GPP per capita and can be used as valid instruments for *Language*.

We undertook similar steps to test for the endogeneity of *Culture* and *Religion*. When *Slavery* is used as an instrument for *Culture*, the null hypothesis for Hausman test statistic is rejected as $\chi^2(2) = 55.48$ and has a *P*-value of 0.000. The Hausman test statistics are 14.91 (*P*-value of 0.0210) and 183.30 (*P*-value of 0.000) for the group 1 and group 2 weather variables respectively. As a result, we estimate a reduced-form equation for each of the instruments. None of the predicted errors are statistically significant at the standard levels of significance. Consequently, we use a larger set of instrumental variables for IVE (Column 3, Table 3) when *Culture* is considered as an endogenous variable. None of the tests provide support for endogeneity for *Religion*. The Hausman test statistic for *Slavery* is 0.22 with a *P*-value of 0.9998, and 5.14 with a *P*-value of 0.5259 for *AvgPrecip* and *AvgRain*.

In summary, for *Language* and *Culture*, the two tests for endogeneity provide evidence of inconsistency of the OLS estimates, but not so for *Religion*. There is no evidence for independent effect on the dependent variable, when *Slavery* and group 2 geography variables are used as instruments and, hence, are considered as valid instruments for both endogenous variables. Group 1 geography variables, that is *AvgTemp* and *AvgTempSq* are valid instruments for *Culture* only. On the basis of these results, we opt for instrumental variable estimation (IVE) as an alternative to OLS to estimate the effects of social diversity on GSP/GPP per capita.

Columns (1–2), Table 3 report the estimates with IVE when only one of the causal variables are treated as endogenous variable. We also use *Slavery* and group 2 geography variables, *AvgPrecip* and *AvgRain*, as instruments. Column (1), Table 3 provides estimates when only *Language* is treated as

¹⁰We retrieve the data from the dataset provided by Basher and Lagerlöf (2006) for public use: <http://www.arts.york.ca/econ/lagerloef/PubDataUSCan.dta>

Table 3. Instrumental Variable Estimation

	ln GSP/GPP per capita				
	(1)	(2)	(3)	(4)	(5)
Constant	8.0031*** (0.2859)	7.7225*** (0.3568)	7.8427*** (0.2883)	7.9360*** (0.3805)	8.0412*** (0.3746)
Language	2.2479*** (0.2943)	2.0891*** (0.27921)	2.0175*** (0.2455)	2.2474*** (0.2949)	
Language 1					2.1792*** (0.2863)
Religion	0.6646*** (0.2454)	0.5657** (0.2574)	0.5845** (0.2476)	0.6466** (0.2561)	0.6289*** (0.2656)
Culture	1.5637*** (0.2360)	1.9753*** (0.3874)	1.8269*** (0.2919)	1.6548*** (0.4456)	1.6098*** (0.4446)
EFS	0.0915*** (0.0217)	0.1022*** (0.0234)	0.0997*** (0.0223)	0.0934*** (0.0235)	0.0897*** (0.0235)
Language * LB	-9.7489*** (1.8058)	-8.1745*** (1.5293)	-8.0015*** (1.4562)	-8.5995*** (1.8264)	
Language 1 * LB					-9.2796*** (1.7258)
Canada_dummy	0.5562*** (0.1193)	0.4557*** (0.1153)	0.4546*** (0.1118)	0.5439*** (0.1298)	0.4930*** (0.1293)
N	58	58	58	58	58
Diagnostics					
R ²	0.6931	0.6698	0.6900	0.6918	0.6899
Adjusted R ²	0.6570	0.6310	0.6536	0.6555	0.6535
Sargan χ^2	0.8188	2.7581	3.1732	0.1481	0.4025
[P-value]	[0.8450]	[0.4304]	[0.6733]	[0.9286]	[0.8177]

Notes: Instruments set used-Column (1)-(2): *Slavery*, *AvgPrecip* and *AvgRain* for only *Language*(Column 1) and only *Culture*(Column 2); Column (3): *Slavery*, *AvgTemp*, *AvgTempSq*, *AvgPrecip* and *AvgRain* for only *Culture*; Column (4)-(5): *Slavery*, *AvgPrecip* and *AvgRain* for both *Language* and *Culture* (Column 4), and *Language 1* and *Culture* (Column 5).

***, ** and * indicate significance level of 1 per cent, 5 per cent and 10 per cent respectively on the basis of two-tailed tests.

Source: Authors' compilation with data from US Bureau of Census and Statistics Canada.

potentially an endogenous variable. We use the *Getsive* method, which follows the *Gets* modelling approach, but utilise IVE instead of OLS. The results are similar to the specific model obtained through OLS (Column 4, Table 2). The estimated coefficients for all three diversity measures are higher with IVE and are statistically significant from zero. More importantly, Column (1), Table 3 supports the hypothesis that social diversity *raises* per capita income and that this positive payoff diminishes, the greater are the linguistic barriers. This is because the coefficient for the interaction term *Language * LB* retains the hypothesised negative sign with a higher estimated impact with IVE in comparison with the final specific model in Column (4), Table 2. The coefficient for *Canada_dummy* retains its positive sign and is statistically significant from zero at the standard levels of significance. This suggests that possibly policy or institutional differences between the two countries may have an important effect on state/provincial levels of economic performance. Overall, the estimated coefficients in Column (1) and Column (2), Table 3 are similar and all of the same sign and levels of significance.

The Sargan test statistic is 0.8188 with the reported *P*-value of 0.8450 for Column (1), Table 3, and is 2.7581 with a *P*-value of 0.4304 for Column (2), Table 3. Thus, we fail to reject the null hypothesis that the error term is uncorrelated with the instruments and conclude that the set of instruments used is valid for each of the endogenous variables.

Column (3), Table 3 reports the IVE estimates when only *Culture* is considered as endogenous with both groups of geography variables as instruments. The estimated coefficients are qualitatively similar to OLS results (Column 4, Table 2), as well as the IVE estimates (Column 1–2, Table 3). The Sargan test statistic provides evidence for validity of the instruments used with a *P*-value of 0.6733 (Column 3, Table 3).

Column (4), Table 3 reports the estimated coefficients when both *Culture* and *Language* are considered as endogenous variables together. The coefficients on the variables of interest, *Language* and *Culture* are positive, the interaction variable with *LB* is negative; and are statistically significant from zero at the 1 per cent level of significance. The estimated coefficient for *Religion* retains its positive sign, and is statistically significant at the 5 per cent level of significance (Column 4, Table 3). In Column (5), Table 3 we check the OLS estimates (Column 5, Table 2) with an alternative interpretation of linguistic diversity, that is *Language 1*. The results are very similar to the previous IVE estimates under different specifications. The Sargan test statistic (Column 4–5, Table 3) provides evidence for validity of the instruments used.

In sum, the IVE results under various specifications with set of instruments used, and endogeneity of casual variables, provide results that are consistent with the OLS estimates. Namely, there is empirical evidence that social diversity *increases* state/provincial GSP/GPP per capita, but this positive economic pay-off is *diminished* by linguistic barriers, as measured by the lack of fluency in the official language.

7. Discussion and Policy Implications

Our results indicate that social diversity is a causal factor contributing to higher levels of per capita income within state and provincial boundaries. However, we also find that linguistic isolation, as measured by inadequate proficiency in the official languages, reduces the positive impact of linguistic diversity in terms of economic pay-offs. These results are robust to alternative model specifications and provide important insights in terms of policy choices in both Canada and the United States, as is the result that, after controlling for other factors, being a Canadian province is associated with higher per capita measure of regional output.

The finding that linguistic barriers reduce the positive pay-off of social diversity could be used to justify raising the minimum English (or French in the case of Quebec) requirement to qualify as an immigrant into the United States or Canada. This may, however, be counterproductive if such a requirement were to reduce the diversity of immigrants. Furthermore, a minimum proficiency in the native language would only affect future migrant flows and would have no impact on the existing population's social diversity. Thus, another option might be to support English

language training of those persons (e.g. family members of skilled migrant or refugees) with inadequate English language skills.

Both the United States and Canada allocate funds to provide subsidised English language training for immigrants under the programmes called English as Second Language (ESL) in the United States, and Language Instruction for Newcomers to Canada (LINC). In 2003, Citizenship and Immigration Canada also launched a programme titled “Enhanced Language Training” to provide job-specific language training skills. However, Chiswick *et al.* (2006) provide evidence that language proficiency of immigrants is closely related with linguistic distance of their mother tongue. Customised programmes with varying degrees of intensity based on the linguistic difference to English (or French in Quebec) may, however, generate better results than the current universal language programme for all groups of immigrants.

Unlike the United States, Canada is officially bilingual and its *Charter of Rights* and the *Constitution Act 1982* provide a limited right to receive publicly funded primary and secondary schooling in either of the two official languages. In practice, however, the province of Quebec has opted out of this section of the Constitution Act, so the implementation of the language rights lacks universal adaptation across the provinces. In Quebec, all immigrants (including Anglophones) are required to send their children to French-language schools. This may enhance the process of social integration in the province, especially for the non-French-speaking immigrant group, but may impede wider connections to other English-speaking provinces.

Our results provide an economic rationale for evaluating multiculturalism or cultural pluralism as immigration policies in settler societies beyond Canada and the United States. Multiculturalism represents a demand for cultural recognition of an immigrant’s culture by both the state and citizens of the country. A potential problem with public funding for multiculturalism is that it may encourage some migrant communities to remain within their existing social groups, thereby creating “silos” rather than “bridges.” This is of particular concern if members of “silo” communities become disconnected from the greater society or country in which they reside. Our results in terms of linguistic barriers provide some support for these concerns and suggest a policy of “inclusive multiculturalism” that encourages social inclusion or a “two-way traffic” in overcoming barriers across social divides.

Social inclusion implies an active role by governments and civil society to establish and maintain intergroup communication. For example, Canada has two specific policies for newly arrived immigrants: the Immigrant Settlement and Adaptation Program (ISAP) and the Host Program (Citizenship and Immigration Canada, 2004). On the basis of needs assessment administered by the service providers, ISAP assists immigration settlement and integration by employing service providers to deliver information on Canadian life, translation and interpretation, referral to community services, solution-focussed counselling and basic employment-related services. The Host Program is a volunteer-based programme whereby newcomers are matched with Canadian volunteers who help them learn about available services in the community, practise their English or French, participate in community activities, understand the Canadian labour market and find job contacts in their field.

The United States has no explicit immigrant integration policies or programmes supported by government (Martin, 2003). Rather, social inclusion largely occurs within civil society. Family members or employers, typically, take principal responsibility for ensuring migrants successfully adapt to the United States and sponsor immigrants. In addition, the US federal government provides grants to private agencies and state and local governments to receive refugees, place them in local communities and provide language training and job counselling and placement services.

The need for social inclusion to maximise the economic benefits of multiculturalism is especially evident for “culturally distant” immigrant communities. Our findings indicate that neither assimilation nor complete acculturation are desirable, but do suggest that pluralistic societies should promote inclusive policies to overcome linguistic and cultural barriers. Although our work is limited to Canada and the United States, we believe it is relevant to other pluralistic societies with large immigrant populations such as Australia and some countries in Europe such as Germany, the United Kingdom and the Netherlands.

8. Concluding Remarks

In this article we investigate if social barriers to communication influence state-provincial GSP and GPP per capita in the United States and Canada. Using data from the 48 contiguous states in the United States and 10 provinces in Canada, we conclude that the impact of social diversity on per capita GSP/GPP is positive and statistically significant, but linguistic isolation reduces the positive effects of diversity. The robustness of our results is investigated and instrumental variables are used to address a possible endogeneity problem.

Despite the necessary caveats in deciphering the subtleties of different social groups, and consequent interactions, our results provide support for policies designed to reduce barriers to communication across social groups within the United States and Canada. We call such an approach “inclusive multiculturalism.” It acknowledges the social and economic benefits of social diversity, but also recognises the need to encourage connections or “bridges” across social divides and groups.

REFERENCES

- Acemoglu, D., Johnson, S. and Robinson, J.A. (2001), ‘The Colonial Origins of Comparative Development’, *The American Economic Review*, **91**, 1369–401.
- Acemoglu, D., Johnson, S. and Robinson, J.A. (2002), ‘Reversal of Fortune: Geography and Institutions in the Making of the Modern World Income Distribution’, *The Quarterly Journal of Economics*, **117**, 1231–94.
- Alesina, A. and Ferrara, E.L. (2002), ‘Who Trusts Others’, *Journal of Public Economics*, **85**, 207–34.
- Alesina, A. and Ferrara, E.L. (2005), ‘Ethnic Diversity and Economic Performance’, *Journal of Economic Literature*, **43** (3), 762–800.
- Alesina, A., Baqir, S. and Easterly, W. (1999), ‘Public Goods and Ethnic Divisions’, *Quarterly Journal of Economics*, **114** (4), 1243–1284.
- Alesina, A., Devleeschauwer, A., Easterly, W., Kurlat, S. and Wacziarg, R. (2003), ‘Fractionalization’, *Journal of Economic Growth*, **8** (2), 155–94.
- Beeson, P. (1987), ‘Total Factor Productivity Growth and Agglomeration Economies in Manufacturing, 1959–1973’, *Journal of Regional Science*, **27** (2), 183–99.
- Bellini, E., Ottaviano, G.I.P., Pinelli, D. and Prarolo, G. (2008), *Cultural Diversity and Economic Performance: Evidence from European Regions*. Hamburg Institute of International Economics, Hamburg.
- Borjas, G.J. (1995), ‘Ethnicity, Neighborhoods, and Human-Capital Externalities’, *The American Economic Review*, **85** (3), 365–90.
- Borjas, G.J. (2003), ‘The Labour Demand Curve is Downward Sloping: Reexamining the Impact of Immigration on the Labour Market’, *Quarterly Journal of Economics*, **118**, 1335–74.
- Chiswick, B.R., Lee, Y.L. and Miller, P.W. (2006), ‘Immigrants’ Language Skills and Visa Category’, *International Migration Review* **40** (2), 419–50.
- Ciccone, A. and Hall, R.E. (1996), ‘Productivity and the Density of Economic Activity’, *American Economic Review*, **86** (1), 54–70.
- Citizenship and Immigration Canada (2004), ‘Evaluation of the Immigration Settlement and Adaptation Program’. Available at: <http://www.cic.gc.ca/english/research/evaluation/isap/exec-summary.html>.
- Collier, P. (2000), ‘Ethnicity, Politics and Economic Performance’, *Economics and Politics*, **12** (3), 225–45.
- Easterly, W. and Levine, R. (1997), ‘Africa’s Growth Tragedy: Policies and Economic Divisions’, *The Quarterly Journal of Economics*, **112** (4), 1203–50.
- Florida, R. (2002), ‘Bohemia and Economic Geography’, *Journal of Economic Geography*, **2** (1), 55–71.
- Fu, S. (2007), ‘Smart Cafe Cities: Testing Human Capital Externalities’, *Journal of Urban Economics*, **61**, 86–111.
- Glaeser, E.L., Kallal, H.D., Scheinkman, J.A. and Shleifer, A. (1992), ‘Growth in Cities’, *Journal of Political Economy*, **100** (6), 1126–52.
- Grafton, R.Q., Knowles, S. and Owen, D.P. (2002), ‘Social Divergence and Productivity: Making a Connection’, in Sharpe A., St-Hilaire F. and Banting K. (eds), *The Review of Economic Performance and Social Progress 2002: Towards an Understanding of Productivity*. The Institute for Research on Public Policy, Montreal; 203–224.
- Grafton, R.Q., Knowles, S. and Owen, D.P. (2004), ‘Total Factor Productivity, Per Capita Income and Social Divergence’, *The Economic Record*, **80** (250), 302–13.
- Grafton, R.Q., Kompas, T. and Owen, D.P. (2007), ‘Bridging the Barriers: Knowledge Connections, Productivity and Capital Accumulation’, *Journal of Productivity Analysis*, **27** (3), 219–31.
- Hendricks, L. (2004), *Why does Educational Attainment Differ Across U.S. States?* Department of Economics, University of Munich, Munich.
- Higgins, M.J., Levy, D. and Young, A.T. (2006), ‘Growth and Convergence Across the United States: Evidence from County-Level Data’, *Review of Economics and Statistics*, **88** (4), 671–81.
- Jacobs, J. (1969), *The Economy of Cities*. Random House, New York.

- Krolzig, H.M. and Hendry, D.F. (2001), 'Computer Automation of General-to-Specific Model Selection Procedures', *Journal of Economic Dynamics & Control*, **25** (6–7), 831–66.
- Kymlicka, W. (2003a), 'Being Canadian', *Government and Opposition*, **38** (3), 357–85.
- Kymlicka, W. (2003b), 'Immigration, Citizenship, Multiculturalism: Exploring the links', *The Political Quarterly*, **74** (s1), 195–208.
- Leigh, A. (2006a), 'Diversity, "Trust and Redistribution"', *Dialogue*, **25** (3), 43–9.
- Leigh, A. (2006b), 'Trust, Inequality and Ethnic Heterogeneity', *Economic Record* **82** (258), 268–80.
- Martin, S. (2003), 'The Politics of US Immigration Reform', *The Political Quarterly*, **74** (S1), 132–49.
- Ottaviano, G.I.P. and Peri, G. (2005), 'Cities and Culture', *Journal of Urban Economics*, **58**, 304–37.
- Peri, G. and Sparber, C. (2009), 'Task Specialization, Immigration, and Wages', *American Economic Journal*, **1**, 135–69.
- Putnam, R.D. (2000), *Bowling Alone: The Collapse and Revival of American Community*. Simon and Schuster, New York.
- Rauch, J.E. (1993), 'Productivity Gains from Geographic Concentration of Human Capital: Evidence from Cities', *Journal of Urban Economics*, **34**, 380–400.
- Rigby, D.L. and Essletzbichler, R. (2002), 'Agglomeration Economies and Productivity Differences in US Cities', *Journal of Economic Geography*, **2** (4), 407–32.
- Rodrik, D., Subramanian, A. and Trebbi, F. (2004), 'Institutions Rule: The Primacy of Institutions Over Geography and Integration in Economic Development', *Journal of Economic Growth*, **9** (2), 131–65.
- Sedgley, N. and Elmslie, B. (2004), 'The Geographic Concentration of Knowledge: Scale, Agglomeration, and Congestion in Innovation Across U.S. States', *International Regional Science Review*, **27** (2), 117–37.
- Shapiro, J.M. (2006), 'Smart Cities: Quality of Life, Productivity and the Growth Effects of Human Capital', *The Review of Economics and Statistics*, **88** (2), 324–35.
- Simon, C.J. and Nardinelli, C. (2002), 'Human Capital and Rise of American Cities, 1900–1990', *Regional Science and Urban Economics*, **32**, 59–96.

Appendix I. Definitions

Ln(GSP/GPP): Gross State Product/Gross Province Product per capita (in natural logs)

Language, Religion, Culture: Fractionalisation indexes for groups identified on the basis of language, race, religion and origin of birth respectively.

Education: Percentage of population with a bachelor degree or higher

Urbanisation: Percentage of state population living in metropolitan areas

PopDensity: Average number of inhabitants per square mile of land area

EFS: Economic Freedom index, measured at the state level

EFF: Economic Freedom index, measured at the federal level

LB: Percentage of linguistically isolated population

Language 1: Linguistic diversity based on broader linguistic roots

Appendix II. Summary Statistics

Variables	Mean	Std. Dev.	Min	Max
<i>Ln(GSP/GPP)</i>	10.3624	0.1827	10.0233	10.8979
<i>Language</i>	0.2246	0.1347	0.0316	0.5578
<i>Religion</i>	0.4423	0.1056	0.0558	0.5584
<i>Culture</i>	0.6648	0.0801	0.3900	0.7616
<i>Education</i>	0.2285	0.0601	0.0953	0.346
<i>Urbanisation</i>	0.7075	0.1486	0.382	0.944
<i>Pop Density</i>	1.7352	0.7086	0.1461	3.0548
<i>EFF</i>	6.75	0.9608	4.3	8.3
<i>EFS</i>	6.33	0.9204	3.8	8.2
<i>LB</i>	0.0669	0.0905	0.002	0.36
<i>Language 1</i>	0.2153	0.1410	0.0317	0.5467

Appendix III. Correlation Matrix for Independent Variables

	<i>Language</i>	<i>Religion</i>	<i>Culture</i>	<i>Education</i>	<i>Urbanisation</i>	<i>EFF</i>	<i>EFS</i>
<i>Language</i>	1.00						
<i>Religion</i>	0.0542	1.00					
<i>Culture</i>	-0.4735	-0.1564	1.00				
<i>Education</i>	0.0964	0.6775	0.0032	1.00			
<i>Urbanisation</i>	0.6825	0.3124	-0.2646	0.3976	1.00		
<i>EFF</i>	-0.0086	0.6733	-0.1277	0.6219	0.3634	1.00	
<i>EFS</i>	-0.1124	0.6524	-0.2129	0.5618	0.2406	0.8843	1.00

Appendix IV. Regions of North America

Regions	States/Provinces
Northeast	Connecticut, Maine, Massachusetts, New Hampshire, Rhode Island, Vermont, New Jersey, New York, Pennsylvania, Newfoundland and Labrador, Nova Scotia, Ontario, Quebec, Prince Edward Island, New Brunswick
Midwest	Indiana, Illinois, Michigan, Ohio, Wisconsin, Iowa, Kansas, Minnesota, Missouri, Nebraska, North Dakota, South Dakota, Manitoba, Saskatchewan, Alberta
South	Delaware, Florida, Georgia, Maryland, North Carolina, South Carolina, Virginia, West Virginia, Alabama, Kentucky, Mississippi, Tennessee, Arkansas, Louisiana, Oklahoma, Texas
West	Arizona, Colorado, Idaho, New Mexico, Montana, Utah, Nevada, Wyoming, California, Oregon, Washington, British Columbia