UBC CENTRE FOR
HEALTH SERVICES AND
POLICY RESEARCH

A Review of Methods for Deriving an Index for Socioeconomic Status in British Columbia

April 2013

Kyle Vincent, PhD
Jason M. Sutherland, PhD

UBC Centre for Health Services and Policy Research
Vancouver, British Columbia

a place of mind

About CHSPR

The Centre for Health Services and Policy Research (CHSPR) is an independent research centre based at the University of British Columbia. CHSPR's mission is to advance scientific enquiry into issues of health in population groups, and ways in which health services can best be organized, funded and delivered. Our researchers carry out a diverse program of applied health services and population health research under this agenda. The Centre's work is:

- Independent
- Population based
- Policy relevant
- Interdisciplinary
- Privacy sensitive

CHSPR aims to contribute to the improvement of population health by ensuring our research is relevant to contemporary health policy concerns and by working closely with decision makers to actively translate research findings into policy options. Our researchers are active participants in many policy-making forums and provide advice and assistance to both government and non-government organizations in British Columbia (BC), Canada and abroad.

Executive Summary

Socioeconomic status (SES) is a measure of the economic and social status of an individual or group of individuals based on education, income, occupation, and other relevant indicators, relative to other members of the population. It is often used to refer to a geographic region's combined economic and social position relative to other areas. In our study the measurement unit of geographic region are the dissemination areas located within the province of British Columbia (BC).

This is a technical report that responds to the BC Ministry of Health's need for an indicator of SES. This report has two main objectives: a review of the development of the Ministry's prototype index for measuring SES and the proposal of recommendations for improving the methods for measuring SES. The former objective is based on reviewing an original analysis (Regional Measures, 2012) prepared for the BC Ministry of Health in 2012, known as the prototype index of SES.

The latter objective is achieved by validating the principal variable approach that was used to develop the prototype index through a review of the evidence and methods (Regional Measures, 2012). Then, the report proposes adjustments to the methods used in the prototype index for measuring SES. A comparison of the results of the two methods for determining the SES index finds that the latter approach is consistent with the methods present in the literature.

This report then proposes an alternative approach to measuring SES that utilizes a two-stage principal components analysis. The benefit of this approach is that it facilitates a bootstrap analysis in order to estimate the variability of the SES scores and range of rankings of the LHAs unavailable using the former approach. The two-stage principal components analysis approach is also consistent with the methods present in the literature.

The final section of this technical report includes recommendations for BC's Ministry of Health for deriving a SES indicator. All analyses in this report were carried out in the R statistical programming language using publicly available data.

Table of Contents

1 Introduction

The British Columbia Ministry of Health (BC MoH) recently developed a *prototype index* for measuring the relative socioeconomic status (SES) of populations in dissemination areas (DA) (and in turn local health areas (LHA) in the province of BC (Regional Measures, 2012). Such measures of SES are relevant to governments as they can reveal which areas of the province are most at risk for economic hardship, declines in education, and/or demand for health care, to name a few.

This technical report accomplishes two objectives. First, it compares and contrasts the methods for determining the aforementioned prototype index (Regional Measures, 2012) for measuring the relative SES of populations in DAs and LHAs in BC to alternative statistical methods. Second, this report then proposes a two-step approach for determining the SES index. This second approach is independent of the first objective and is intended to be a comparative approach to the first method and the approach is used due to its ability to derive variance estimates and confidence intervals for the SES index scores (and rankings) of the LHAs.

The report is organized as follows. Section 2 gives a brief overview of the subset of the 2006 census data set that was used in our study and highlights missing census data from certain DAs. Section 3 gives an overview of some of the more well-known SES indices that are present in the literature, as well as some of the primary concerns with the existing SES indices, and then outlines the primary concerns found in the review of the development of the BC MoH's prototype index. The motivation for the development of our two SES indices is also provided. Section 4 presents an alternative SES index that we have developed, namely one which is a variation of the BC MoH's prototype index based on a method of principal components analysis (PCA) called the principal variable approach. Section 5 presents our second method for developing an SES index, namely one which is based on a two-stage PCA and that contains a bootstrap analysis to obtain confidence intervals for the SES scores of the LHAs. Section 6 then provides a discussion of the study with recommendations for the BC MoH. We also provide suggestions for the future use and validation of the SES indices that are developed in this report.

We have uploaded the original data set (and variations of this data set) that was used in our study, as well as the R code that was developed for the analyses. The material can be found at www.healthcarefunding.ca. These data sets and R code can potentially serve as a foundation for future PCA-type analyses that are of interest to the BC MoH.

2 Data

The data that we use in our study is based on the 2006 Canadian census data. An overview of the data set can be found at the following web link:

http://www.bcstats.gov.bc.ca/StatisticsBySubject/Census/OpenData.aspx

Overview of Our Data Set

Our data set is comprised of 78 variables, where each variable is classified a priori to one of six domains: education, income, employment, housing, social, and occupation. Each variable was classified by the BC MoH as belonging to one of the domains, and this will facilitate the interpretation of the SES index.

For all analyses discussed in this report, the basis was the "Complete Reduced Database Final" excel spreadsheet, as outlined in Regional Measures (2012). A dissemination area (DA) is a census geographic unit of Canada, there are 7,471 in total, but many do not have associated data. In the data set used for this study there were a total of 6,440 DAs with data (which are the unit of measurement in this study) each with a population size of 400-700 residents.

A local health area (LHA) is an aggregation of DAs that makes up a contiguous region of BC. Regional Measures (2012) reveals that the 6,440 DAs were distributed amongst 78 LHAs. As there are more than 78 LHAs in BC this gave rise to some primary concerns with the assignment of DAs to LHAs as well as which LHAs appeared to be missing. For example, it was found in this original data set that the Westside LHA was originally comprised of the following six LHAs: Westside, Downtown Eastside, South Vancouver, Northeast, City Centre, and Midtown.

The translation file developed by BC STAT and obtained by the BC MoH was used to update the assignments of DAs to LHAs. It was then found that a total of 89 LHAs were available for consideration of ranking in our study. Three LHAs, namely Central Coast, Nisga'a, and Telegraph Creek, were not affiliated with any of the 6,440 DAs in our data set and therefore are not provided a ranking in our study.

This project is limited by the use of the 2006 BC census data. Therefore, it is suggested that for future validation of our methods for developing SES indices, our indices should be validated against existing SES indices *with the census/national data that was used in the respective development* of the existing indices in order to determine the robustness of our methods. Furthermore, more recent BC census data may generate results that differ significantly from these findings as some of the census data may have changed due to migration and/or changes in assignments of DAs to LHAs. Therefore, it is strongly suggested that the analyses outlined in our study be reapplied to the most recent data when it becomes available.

For further information on the data set used, please see the Data section in Regional Measures (2012).

Dissemination Areas with Missing Data

A list of the 7,471 DAs of BC was obtained from the BC Ministry of Health. Population counts that are publicly available from Statistics Canada were obtained for these DAs (Census of Canada, 2006). In the data provided by the BC MoH, a total of 6,440 DAs had census data available for analysis. In the first step of our review of the data, we investigated the population counts and geographic locations of the 7,471-6,440=1,031 DAs that were missing census data.

Figure 1 provides an overview of the DAs in BC, highlighted by those which were included or excluded in our data set; those areas highlighted in brown are those DAs that are included in our data set and the areas in pink represent those DAs that were not included in our data set. The census subdivisions outlined in red represent First Nations reserves. Notice that DAs affiliated with First Nations reserves are excluded from our data set at a higher rate, perhaps due to the increase in Aboriginals living off reserve and in urban centers in recent years (Aboriginal Peoples in Canada in 2006: Inuit, Métis and First Nations, 2006 Census). See section A.1 of the Appendix for enlarged portions of the map.

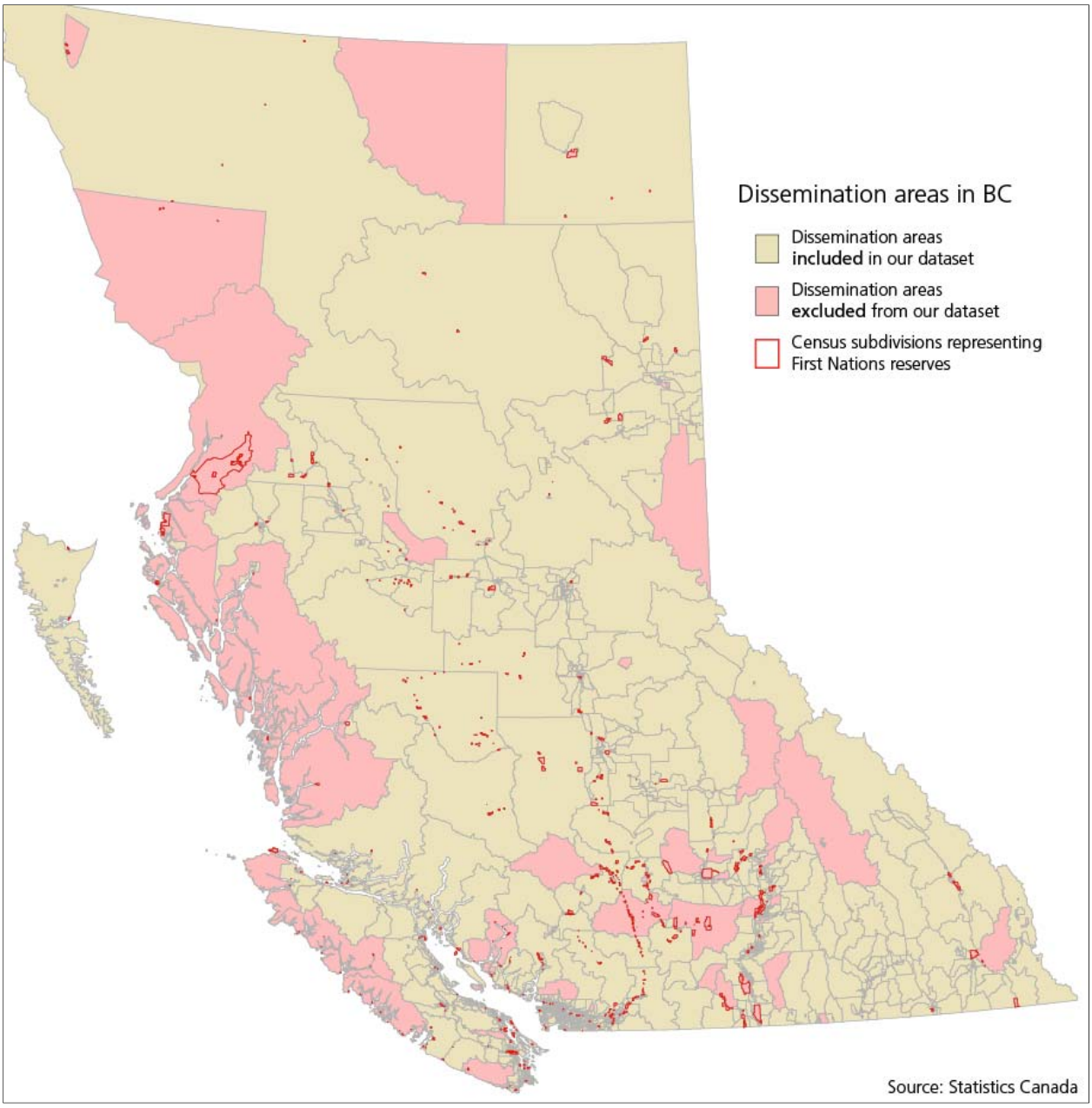

Figure 1: An overview of the DAs in BC. Areas highlighted in brown are those DAs that are included in our data set and the areas in pink represent those DAs that were not included in our data set. The areas outlined in red represent First Nations reserves.

An examination of the population counts of the 1,031 excluded DAs revealed that a large number had population counts that were much lower than those that were included in our data set. Moreover, 248 of the 1,031 excluded DAs had zero population counts.

It was also found that some excluded DAs had recorded relatively large population counts but were not included in our data set (for example in the Vancouver-Westside LHA). This issue calls for

further review from the BC MoH as we are unclear as to why such an issue would present itself in the data set.

We have included a table that records the counts of the total number of DAs and the number in our data set, by LHA, as well as the corresponding population counts, by LHA, in section A.1 of the Appendix. This table can help to clarify some of the expected bias in the results as scores may be based on a small and/or unrepresentative subset of the population of LHAs. For example, in our analysis it was found that the Bella Coola Valley, Stikine, and Upper Skeena LHAs only included 46%, 44% and 35% of the population census data, respectively. As revealed in Figure 1, the data from such LHAs are more likely to come from urban areas. This will likely deflate the SES score for the LHAs as recent studies have shown that rural areas are more likely to be deprived (Human Early Learning Partnership, 2009), therefore giving a better than expected ranking.

3 Overview of Methods

There has been an interest in the use of principal components analysis (PCA) to develop SES indices. For example, Salmond and Crampton (2002) developed the New Zealand (NZDep) index where a PCA is applied to a set of preselected variables from the New Zealand census data and the index weights are based on the *first principal component*. Similarly, Messer et al. (2006) created an index based on the first principal component of a PCA that is applied to American census data based on a set of preselected variables. Pampalon and Raymond (2000) also develop the Institut national de santé publique du Québec (INSPQ) index based on the first two principal components of a PCA that is applied to census data from Québec. In their study, the first principal component comprises the weights for a "material-based" deprivation index and the second principal component comprises the weights for a "social-based" deprivation index. In the literature it is common to see the term *deprivation index* used in place of *socioeconomic status index*, and hence we may use these terms interchangeably in our report. For further information and review of some of the recent methods used to develop SES indices, see Regional Measures (2012).

Currently, two SES indices have been developed for and applied to BC census data: the Overall Regional Socio-Economic Index (BC Stats, 2011) and the Vancouver Area Neighbourhood Deprivation Index (Bell et al., 2008). The Overall Regional Socio-Economic Index (ORSEI) is based on a composite index of six separate indices, where each index is based on some form of regional hardship and is a composite of three or four variables. For example, the Index of Crime is based on the number of serious property crimes, violent crime offences and total number of serious crimes, per 1,000 persons living in a regional district or local health authority. The weighting of the variables to create the score within each index is based on researcher defined weights and due to the high level of subjectivity in the development of the ORSEI, this method for measuring SES has recently come under scrutiny (see Regional Measures (2012) Appendix 1). Similarly, the choice of variables and the weighting they receive in the Vancouver Area Neighbourhood Deprivation Index (VANDIX) was based on consultations with Medical Health Officers in the province and is based almost entirely on a subjective scale for measuring SES. Neither the ORSEI nor the VANDIX have an element of PCA in their analysis. We have addressed this gap by developing an empirically-derived (using PCA) SES index for BC.

Throughout this report, we assume that the reader is familiar with principal components and bootstrap methods. These methods are commonly applied to a range of methodological problems and this report provides only an accessible summary of the methods; the reader is advised to consult other sources for a thorough and rigorous introduction to principal components or bootstrap methodologies.

3.1 Analysis of Prototype SES Index Methods

The approach we have taken for developing an SES index for use in BC relies on the use of a principal components analysis to reduce the large number of variables contained within the 2006 Canadian census data set to a relatively small number of principal components to work with. In effect, PCA reduces the dimensionality of a large number of variables without severely compromising the amount of information contained within the variables.

Principal components analysis is a method for transforming data into new values that are linearly uncorrelated. These new and uncorrelated data (which are representations of the original data) are called principal components. Thus, principal components are ordered vectors that contain information from the original data set. Principal components are ranked in descending order according to the variance associated with each vector (commonly referred to as an *eigenvalue*). The more variable the first few principal components are, the more information they contain (Joliffe, 2002) and are therefore "representative" of the most influential behaviour of the full data set. The analytic strategy using a PCA-based approach is to utilize the information contained within the first few principal components to represent the variability in the original data.

The recommended approach that was used for developing the BC MoH prototype index was based on a specific PCA called the *Principal Variable model*. This method is generally well-received and several leading researchers in the field of PCA have proposed using the principal variable method (for example, see Joliffe (2002) and McCabe (1984)). With the Principal Variable approach an intial PCA is performed on the data set and, via the observation of the magnitude of the variances of the ordered principal components, it is then determined by the analyst how many of the original variables from the data set should be retained for the next PCA. From here, a subset of the variables are selected according to some chosen criteria (Joliffe, 2002) and then a second PCA is run on the subset of variables. The final index is based on the first principal component from this PCA.

Implementation of the methods used to design the prototype SES Index

Although the results based on the prototype index indicate that this approach is promising, some concerns became apparent when attempting to recreate the index (that is, this is our review of the Ministry's prototype SES index).

First, there appears to have been a miscount in the number of variables that are removed from the data set in the data reduction stage. Also, there is no discussion of the method used to determine which variables should be removed based on their weights. Therefore, the method used to decide upon the fourty variables (Regional Measures, 2012, Table A5.1) to be used for the principal variable stage of the analysis is unknown.

Second, the principal variable model is applied but there is no discussion regarding the criteria applied for retaining the 12 variables (Regional Measures, 2012 Table A5.2) from the subset of the 40 variables identified in the previous step. Joliffe (2002) has outlined four common criteria that

are used in such an approach and although each one, in theory, can be attempted to determine if the results coincide with those presented in the report, however, this was not pursued. The omission of the discussion on methods has implications for evaluating the sensitivity of the findings subject to the established criteria.

Third, the author of the Ministry's report regarding the construction of the prototype index expresses concern about an over representation of the education variables that are used in the final step of the principal components analysis (seven of the 12 variables used in the 12 principal variable model are education-based; Regional Measures, 2012 Table A5.2). The over representation of education variables may lead to the index being less well-received amongst the BC MoH as compared to indices like the ORSEI (BC Stats, 2011) and the VANDIX (Bell et al. (2008)). No mention of attempts to resolve this potential issue is made by the author of Regional Measures (2012).

Validation of the prototype index against some of the existing methods has done well (for example, the INSPQ (2008), ORSEI (BC Stats, 2011) and VANDIX (Bell et al., 2008). As PCA allows for many variants of analyses this provided the motivation to explore the use of other PCA approaches to determine if any give rise to a complementary index that is more valid than the prototype index.

4 Socioeconomic Index 1: Principal Variable Approach

We applied specified criteria and methods to derive the *complementary SES index* using the updated data set. The steps and results from the PCA are outlined in the following subsections.

4.1 Data Reduction for Complementary SES Index

We performed a preliminary analysis that omits data reduction steps to determine the benefit of an immediate data reduction in the analysis (as opposed to a two-phase approach used in the development of the prototype indicator). This preliminary analysis revealed that a PCA with an immediate data reduction step yields an SES index that avoids over-weighting (a combination) of some variables. For example, there are many different measures of education, though only a few may be required at the principal variable stage of the analysis to measure how differing levels of education contribute to the distribution of deprivation across DAs.

We found that there was a large departure from normality in the final scores with the principal variable method and this resulted in a large number of extreme LHA scores. This finding suggests that this approach may not be appropriate since it will yield inplausible results.

On the basis, we have decided to incorporate a data reduction step (as was done in the construction of the prototype index) into the analysis and to include a test for normality in the distribution of the final SES scores amongst the DAs.

Commencing with the 78 variables that are present in the “Data Working Version" file (we utilized the unstandardized raw data at this step), we first removed the count variables and those with a median of zero. We are in agreement with Regional Measures (2012) that count variables do not provide insight into the SES of the DAs (it can be argued that count variables cannot even serve as weak proxies to the SES status of DAs) and that variables with a median of zero could potentially skew the final scores of the DAs (as all observations are greater than or equal to zero). A total of 19 variables were removed; 2 count variables and 17 variables that had a median of zero. A total of 59 variables remained after this step.

Our analytic objective was to find variables that tend to be uncorrelated. The rationale for including less correlated variables is to include those that impart potentially additional information regarding SES, plus it reduces the number of redundant variables.

To do so, we considered the correlation matrix of the variables and following the approach in Regional Measures (2012), we sequentially removed variables that had a Spearman rank correlation whose absolute value exceeded 0.8 (the use of other values, like 0.7 and 0.9, gave rise to final SES scores for the DAs that departed significantly from normality). We considered the list of the remaining variables in the “Data Working Version” file, and first compared the first variable

to each of the remaining variables. If the paired measures exceeded 0.8 then the variable succeeding the first variable was removed. Next we compared the second (remaining) variable to each of the succeeding remaining variables and again, if the paired measure exceeded 0.8 then we removed the succeeding variable. We did this sequentially until we reached the end of the list and found that a total of 16 variables were removed and a total of 43 variables remained. Histograms of the distributions of the variables were obtained and it was found that there were few outliers in each of the variables' distributions.

Finally, we scaled our data so that the variables exhibited a mean of zero and a standard deviation of one (this was appropriate as none of the variables were ordinal or categorical). Using a PCA that rests on the use of the correlation matrix of the standardized data, as opposed to the raw data, has been advocated by many researchers in the field of PCA (for example, see Joliffe (2002) and Jackson (1991). This approach is then based on variables that share the same units of measurement and therefore will avoid a potential overweighing of a subset of the variables.

4.2 The Principal Component Analysis

We ran a principal components analysis on the remaining data set of 43 variables, after scaling the data to work with the correlation matrix in the PCA, to determine how many variables should be retained for the next PCA. Similar to the construction of the prototype index, it was decided that the number of variables to retain was to be equivalent to the number of principal components whose variance exceeds the value of one. Taking those components that explain a higher percentage of the variation relative to the raw data measurement scale gives an intuitive reason to retain the number of variables equal to the number of components that exceed this threshold. This resulted in 12 principal components that had a corresponding variance greater than one. Figure 2 plots the observed variances of the principal components from this step.

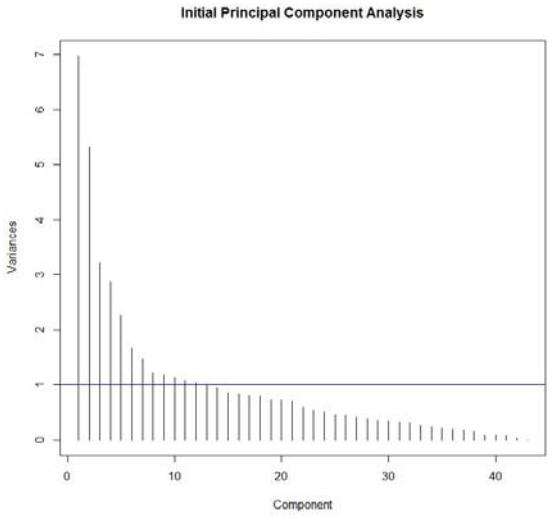

Figure 2: A plot of the variances of the principal components. The horizontal line indicates our threshold to decide how many variables to retain; the first 12 principal components have a variance greater than one.

We carried out the principal variable step using a criterion outlined by McCabe (1984). Section A.2 of the Appendix provides the mathematical details on the principal variable step as well as McCabe's (1984) selection process that we used to select 12 variables that were retained for the second PCA.

A second PCA was run on the 12 variables that were retained. Table 1 presents the final 12 variables with the first two loadings from this principal components analysis.

Variable	PC 1	PC 2	Domain
Highest level of educational attainment ages 15-24: Below HS	0.085	0.374	Education
Highest level of educational attainment ages 15-24: College (no univ. Degree)	0.011	-0.241	Education
Highest level of educational attainment ages 25-64: College (no univ. Degree)	-0.096	0.142	Education
Highest level of educational attainment ages 25-64: Bachelor's degree	-0.352	-0.414	Education
Highest level of educational attainment ages 25-64: Post-bachelor's degree/diploma	-0.387	-0.346	Education
Highest level of educational attainment ages 65+: Below HS	0.374	0.093	Education
Unemployment rate, all ages and sexes	0.324	-0.041	Employment
Participation rate, all ages and sexes	-0.206	0.075	Employment
Participation rate, aged 15-24 and all sexes	-0.007	0.015	Employment
Median Income, all ages and sexes	-0.515	0.066	Income
Percent of occupied dwellings that are tenant-occupied	0.300	-0.497	Housing
Proportion of households with one occupant	0.251	-0.470	Social

Table 1: Constructing the complementary index: The 12 variables that are retained for the principal variable approach with the loadings from the first two principal components.

The magnitude and signs of the first principal component reveals the underlying nature of the weights that are used for the complementary index. For example, education variables based on higher degrees receive a negative score, indicating that the higher the presence of an educated population then the lower, and therefore better, the SES (deprivation) score. Similarly, the single social variable receives a positive score, indicating that the greater the number of individuals living alone the higher the SES (deprivation) score.

The second principal component is not as easily interpreted as the first. This may be due to the orthogonality of these two principal components (Joliffe, 2002). For example, if an SES index was based primarily on the second principal component then the more education amongst the population and/or the greater the number of occupants living alone and in tenant occupied dwellings, the lower the score.

Relative to the scores based on the first principal component for the prototype index, the six domains are better represented in the scores from the first principal component based on our approach. Every domain now contributes at least one variable to the first principal component (the

housing variable was absent from this part of the analysis with the prototype index; Regional Measures, 2012, page 20). As was found with the construction of the prototype index, a potential over-representation of the education variables is evident. However, this issue is resolved with the use of the two-stage PCA index we have developed and is introduced in the next section.

Utilizing the Second Principal Component

Regional Measures (2012) acknowledged the potential use of a second principal component in the development of an SES index. Previous studies (for example, Pampalon and Raymond (2000)) have advocated the use of the first two principal components when developing an SES index (or, as Pampalon and Raymond refer to as a deprivation index). In their study, Pampalon and Raymond utilized six preselected variables for a PCA; three variables were "material-based" (one from each of the income, education, and employment domains) and three variables were "social-based" (two from the social domain and one from the housing domain). After performing the PCA, the first two principal components assigned weights in such a manner that a material deprivation index could be based on the first principal component and a social deprivation index could be based on the second principal component. This was possible because the three variables of interest were assigned the majority of weight for the corresponding principal component. In our study, the magnitude of the scores from the first principal component reveal that both material-based and social-based behaviour is semi-accounted for as there are variables from each domain that make up the first principal component. Taking a sum of the first two principal components will place a disproportionate amount of weight on education and, by the signs of the weights on the tenant occupancy and living alone variables, their influence on the final scores will be negligible. Therefore, making use of the second principal component is not suggested for this analysis.

4.3 Results for the Complementary Index

Figure 3 presents a histogram of the SES scores (according to DA) that are based on the use of the complementary index described above. The results from a Jarque-Bera test for normality are also given in Table 2. However, since the sample size is very large ($n = 6{,}440$), a qq-plot to subjectively test for normality is provided in Figure 4.

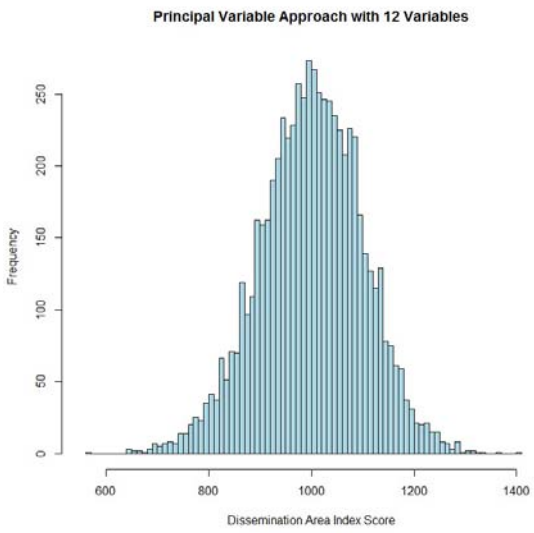

Figure 3: Complementary index: The DA scores. The DA receiving the highest SES score was DA 59153183 (from the Downtown- Eastside), receiving an SES score of 1403.2. The DA receiving the lowest SES score was DA 59150020 (from the West Vancouver-Bowen Island LHA), receiving an SES score of 563.7.

χ -squared = 27.2265	df = 2	p-value = 1.224e-06

Table 2: Complementary index. Jarque-Bera test for normality of the DA deprivation scores.

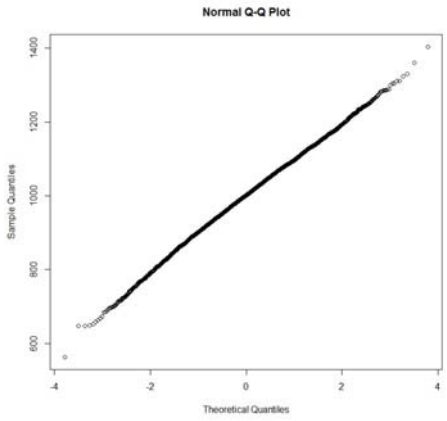

Figure 4: Complementary index. A qq-plot of the DA scores.

The qq-plot reveals a straight line and that no serious departure from normality is detected. We conclude that a suitable choice of method based on a PCA has been chosen to reveal those areas that are most deprived without ambiguity in the final scores that skewed distributions may show at the extremes. For example, skewed distributions can have a high concentration of SES scores at the upper and/or lower range of the distribution and this could give rise to a large number of areas that appear to be approximately equally deprived.

Tables 3 and 4 provide the five highest (most deprived) and lowest (least deprived) SES scores for DAs by complementary index, respectively. SES scores based on the prototype index for the DAs are also provided. The Spearman correlation of the SES scores of the DAs based on the two indices came out strong, and therefore the close agreement in scores when considering the most deprived DAs was to be anticipated.

Dissemination Area (LHA)	Complementary Index Score	Prototype Index Score
59153183 (Downtown Eastside)	1403.2	1431.7
59150350 (Downtown Eastside)	1360.9	1376.5
59330063 (Kamloops)	1330.8	1325.9
59153582 (Downtown Eastside)	1323.4	1324.5
59470072 (Prince Rupert)	1310.8	1280.5

Table 3: The top five SES complementary index scores by DA.

Dissemination Area (LHA)	Complementary Index Score	Prototype Index Score
59150020 (West Vancouver)	563.7	1068.3
59150013 (West Vancouver)	647.0	1003.1
59150018 (West Vancouver)	647.4	1002.9
59150663 (Westside)	649.0	1053.0
59150129 (North Vancouver)	653.0	758.7

Table 4: The bottom five SES complementary index scores by DA.

Tables 5 and 6 provide the five highest (most deprived) and lowest (least deprived) SES scores for LHAs by complementary index, respectively. SES scores based on the prototype index for the LHAs are also provided. Similar to the scores found for the DAs, there is close agreement between the complementary and prototype index when considering the most deprived LHAs. Recall that in the original data set the Westside was comprised of six LHAs, namely the Westside, Downtown Eastside, South Vancouver, Northeast, City Centre, and Midtown, and the score from the prototype index was based on treating all of these LHAs as one unit. In addition, the Stikine LHA was not affiliated with any DAs in the original data set and therefore does not receive a score under the prototype index.

Local Health Area	Complementary Index Score	Prototype Index Score
Upper Skeena	1163.3	1141.5
Keremeos	1130.8	1099.1
Princeton	1117.5	1081.0
North Thompson	1116.5	1071.4
Lake Cowichan	1114.8	1050.7

Table 5: The top five SES complementary index scores by LHA.

Local Health Area	Complementary Index Score	Prototype Index Score
West-Vancouver	837.5	905.1
Westside	852.9	1059.7
North Vancouver	901.6	938.2
Saanich	917.0	949.1
Stikine	919.7	N/A

Table 6: The bottom five SES complementary index scores by LHA.

Section A.3 of the Appendix provides a summary figure of the SES scores and gives the ascending average scores of dissemination area by the 89 LHAs when using the complementary index. Recall that the LHAs Central Coast, Nisga'a, and Telegraph Creek were not affiliated with any of the 6,440 DAs in our updated data set and therefore do not receive a score in the rankings. Both the rankings and the scores are provided to compare and contrast the relative degrees of deprivation of the LHAs.

4.4 Validation of the Complementary Index

We compare our complementary index with the prototype index by considering the SES scores of the DAs based on the two indices. We utilize the scores of the DAs that are based on the use of other common SES indices and the Spearman rank correlation. The Spearman rank correlation is a popular statistic that can be used to robustly measure the correlation between two random variables. In our study we employ the use of the Spearman rank correlation statistic to compare our complementary index with the prototype index via paired comparisons of each over correlations with proxy variables and SES indices that are found in the existing literature. If one of the complementary or prototype index correlates better with one of the variables or SES indices that we consider, then we will say that this index "validates" better than the other with respect to this proxy variable or SES index.

4.4.1 Validation over six proxy variables

Table 7 gives the Spearman rank correlation of the scores from the complementary and prototype indices with one variable taken from each of the six domains, as was done in the validation section of the prototype index (Regional Measures, 2012). These six variables are: the proportion of the population aged 25-64 without high school education (EDNO25: Education), unemployment rate

(UNEMP: Employment), average income per person (INCAV: Income), average value of owner-occupied dwellings (AVVAL: Housing), the proportion of population employed in finance, management, professional and scientific industries (OCCFM: Occupation), and the proportion of families with a lone female parent (LONEPF: Social). Each variable was chosen because it can serve as a weak proxy to the SES of DAs. This serves as a good starting point for validation of the complementary index. The size of the correlation indicates a higher agreement in rank via the variable and the SES index score. The sign of the correlation gives an indication as to the direction of deprivation that each variable contributes towards. For example, the higher the average income per person, the lower (and better) the deprivation score.

Index	EDNO25	UNEMP	INCAV	AVVAL	OCCFM	LONEPF
Complementary	0.636	0.454	-0.776	-0.500	-0.455	0.289
Prototype	0.431	0.420	-0.630	-0.340	-0.216	0.292

Table 7: Spearman rank correlation between SES indices and six variables of interest.

In each case but the last, the complementary index has validated better than the prototype index. It was emphasised in Regional Measures (2012) that the first three variables are commonly used in measures of SES, and our index has validated better in each of the three cases. This is a good indication that our SES index is scoring the DAs in an intuitive manner and which can be easily interpreted.

4.4.2 Validation over the INSPQ

The INSPQ (2008) method for scoring DAs was also considered for validation. We utilized the "Copy of EquivalenceTableCanada2006_A" data that was present in the INSPQ file that was made available by the BC MoH. Of the 6,440 DAs that were present in our data base, it was found that a total of 6,318 DAs received an INSQP score under the "material-based" and "social-based" deprivation index heading (recall that these are based on the use of the first and second principal components of the INSPQ method). We compared the SES scores for these DAs from the complementary index and the prototype index with the INSPQ deprivation scores. Table 8 gives the correlation of the SES scores for the complementary and prototype index with the INSPQ material and social deprivation scores.

Index	INSPQ, material	INSPQ, social
Complementary	0.716	0.343
Prototype	0.537	0.425

Table 8: Spearman rank correlation between SES indices and the first two components of the INSPQ.

We found that the complementary SES index scores validated significantly better with the "material-based" deprivation scores and slightly weaker with the "social-based" deprivation scores that are based on the INSPQ index, relative to those with the prototype index.

Given the magnitude of these correlation scores, we choose to base the complementary index solely on the first principal component, since the scores indicate that the first principal component captures the behavior of both the material and social based aspects of a DA's SES.

4.4.3 Validation over the SEFI

The SEFI (1996) is a widely used SES index and is based on the use of the following variables:

1) The ratio of the population aged 65 or older in a region to the population aged 15-64,

2) Single parent households,

3) Female single parent households,

4) Labour force participation female,

5) Unemployment 15-24, 25-34, 35-44, 45-54, and

6) Education 25-34, 35-44, 45-54.

Our census data does not include specific information for age groups 35-44 and 45-54. We have constructed a pseudo-SEFI index based on available data. In our approach, we utilize the following six variables:

1) The proportion of the population aged 65 or older,

2) Single parent households,

3) Female single parent households,

4) Unemployment rate of females,

5) Unemployment rate of those aged 15+, and

6) Highest level of educational attainment ages 25-64: below high school.

The SEFI proceeds by running a principal component analysis and then basing weights for the final scores on the first principal component.

We found that the Spearman correlation for the SES scores of the DAs based on the complementary index and the SEFI-like index (when applied to our data set) was 0.441. We also found that the Spearman correlation of the SES scores of the DAs obtained with the prototype index and those obtained with the SEFI-like index scored lower at 0.404. With the pseudo-SEFI-like index, the complementary index validates better than the prototype index.

4.4.4 Further Validation

At this time we are unable to validate our index with the ORSEI and the VANDIX indices. However, we found that there was a strong Spearman correlation of 0.79 between the complementary index and the prototype index. We found that the prototype index correlated well with the ORSEI and VANDIX (correlation scores of 0.52 and 0.69, respectively) and that the complementary index has validated better in almost all cases relative to the prototype index in our study. Therefore, validation of the complementary index with respect to the ORSEI and VANDIX is expected to do well.

5 Socioeconomic Index 2: The Two-Stage Principal Components Analysis Approach

The analysis based on the prototype index consists of a computationally expensive step that selects the variables to use (through the Principal Variable model) for the second PCA. These selected variables can yield an index that places a large weight on some of the domains while placing little to no weight on other domains. We have opted for a strategy that is computationally less cumbersome and that avoids this over-weighting issue.

We first reduce the data via a PCA on the variables associated with each of the six domains. As this strategy is much less computationally expensive, it lends itself well to the incorporation of a bootstrap analysis (Efron and Tibshirani, 1993) to derive confidence intervals of the final scores of the LHAs.

In this section we outline the steps and results obtained with the index based on this two-stage principal components analysis approach which we will refer to as the *two-stage index*. We also outline the bootstrap analysis that is used to derive confidence intervals for the scores/rankings of the LHAs.

5.1 Stage 1 of the Principal Component Analysis

After scaling the data to work with the correlation matrix, the first step of the two-stage principal components analysis approach was to run a PCA on each of the subsets of variables that correspond with the six domains (we do not include any sort of immediate data reduction step as we did with the development of the complementary index). We retain the first principal component from each of the six domains as these explained a significant amount of the variation in each of the subsets of the data, shown in Figure 5.

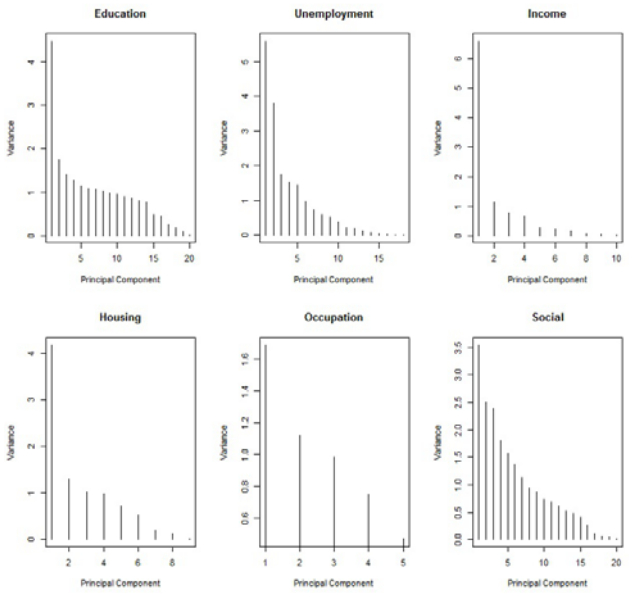

Figure 5: Variances of each of the principal components based on a principal components analysis for each of the six domains.

5.2 Stage 2 of the Principal Components Analysis

The next step was to run a principal components analysis on the first principal components that correspond with the six domains. Each one of these principal components is a vector of length 6,440 and serves as a proxy of information that the corresponding domain contributes to the full data set.

Table 9 provides the loadings from the PCA performed on the set of principal components. These loadings will comprise the weights used for the final index.

Education	Unemployment	Income	Housing	Occupation	Social
-0.452	0.272	0.506	0.322	0.422	-0.426

Table 9: Constructing the two-stage index: Loadings from the first principal component based on a PCA for each category.

The weights shown in Table 9 give rise to an interesting structure, with most weight resting on the education and income variables. It appears at first that DAs with higher income will give rise to a larger SES score. However, there is a large correlation between the education and income variables. It is possible that the weight associated with income is offsetting the large contribution from the weight associated with education.

5.3 Results based on the Two-Stage Index

Figure 6 presents a histogram of the SES scores for the DAs based on the two-stage index. The results from a Jarque-Bera test for normality and a qq-plot to subjectively test for normality are also provided in Table 10 and Figure 7, respectively.

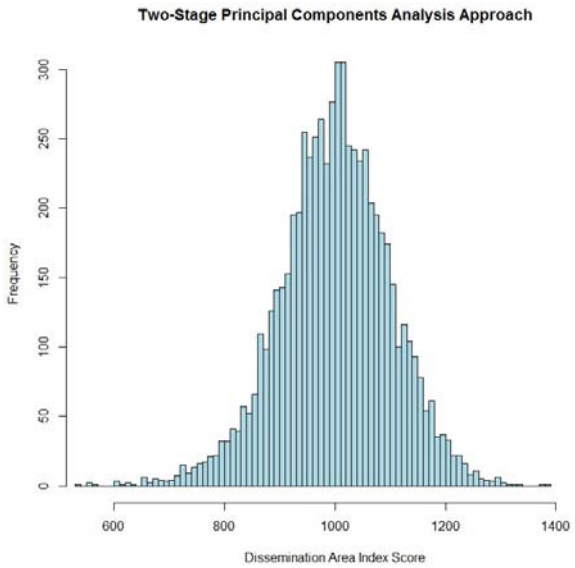

Figure 6: Two-stage index: the DA scores. The DA with the highest SES score was 59090097 (from the Abbotsford LHA), receiving an SES score of 1388.5. The DA with the lowest SES score was 59150008 (from the West Vancouver-Bowen Island LHA), receiving an SES score of 530.1.

χ -squared = 156.7635	df = 2	p-value < 2.2e-16

Table 10: Two-stage index: Jarque-Bera test for normality of the DA deprivation scores.

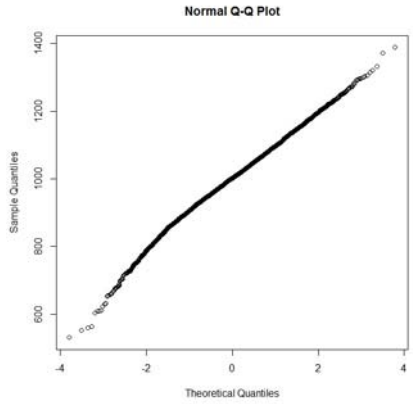

Figure 7: Two-stage index: A qq-plot of the DA scores.

There is a small departure from normality in the scores. However, a straight line in the qq-plot is evident indicating that no serious departure from a normal distribution has occurred.

Tables 11 and 12 provide the five highest (most deprived) and lowest (least deprived) SES scores for DAs by the two-stage index. SES scores based on the prototype index for the DAs are also provided. Similar to the findings with the complementary index, we found that the Spearman

correlation of the SES scores of the DAs based on the two indices came out strong. In this case, it is the scores from the least deprived DAs that are in closer agreement.

Dissemination Area (LHA)	Two-stage Index Score	Prototype Index Score
59090097 (Abbotsford)	1388.5	1253.9
59370193 (Vernon)	1372.9	1229.0
59190229 (Cowichan)	1332.1	1143.5
59070213 (Penticton)	1320.9	1215.0
59090110 (Abbotsford)	1315.1	1219.8

Table 11: The top five SES two-stage index scores by DA.

Dissemination Area (LHA)	Two-stage Index Score	Prototype Index Score
59150008 (West Vancouver)	530.1	811.4
59150954 (Westside)	552.7	738.3
59150953 (Westside)	559.2	813.7
59150662 (Westside)	562.8	546.7
59150032 (West Vancouver)	604.0	836.6

Table 12: The bottom five SES two-stage index scores by DA.

Tables 13 and 14 provide the five highest (most deprived) and lowest (least deprived) SES scores for LHAs (by the two-stage index). SES scores based on the prototype index for the LHAs are also provided. The high correlation of the SES scores for the DAs based on the two-stage and prototype index indicates that close scores for the LHAs should be expected. Recall that the Westside LHA was comprised of six LHAs in the original data set.

Local Health Area	Two-stage Index Score	Prototype Index Score
Keremeos	1176.3	1099.1
South Cariboo	1139.1	1075.5
Kettle Valley	1137.8	1051.4
Princeton	1124.3	1081.0
Creston	1121.3	1053.8

Table 13: The top five five SES two-stage index scores by LHA.

Local Health Area	Two-stage Index Score	Prototype Index Score
West Vancouver-Bowen Island	814.8	905.1
Westside	840.0	1059.7
North Vancouver	899.1	938.2
Coquitlam	942.7	960.0
Delta	943.3	933.7

Table 14: The bottom five SES two-stage index scores by LHA.

Section A.3 of the Appendix provides a summary figure of the SES scores and gives the ascending average scores of DA by the 89 LHAs when using the two-stage index. Recall that the LHAs Central Coast, Nisga'a, and Telegraph Creek were not affiliated with any of the 6,440 DAs in our updated data set and therefore do not receive a score in the rankings. Both the rankings and the scores are provided to compare and contrast the relative degrees of deprivation of the LHAs.

5.4 Validation of the Two-Stage Index

In this subsection we validate the two-stage index in the same manner as the complementary index. We make use of the same SES indices that are present in the literature and the Spearman correlation measure to compare the performance of the two-stage index with the prototype index. We refer the reader to the validation subsection of Section 4 for further details of the validation methods.

5.4.1 Validation over six proxy variables

Table 15 gives the Spearman rank correlation of the SES dissemination scores obtained with the two-stage index approach and a variable that was taken from each of the six domains.

Index	EDNO25	UNEMP	INCAV	AVVAL	OCCFM	LONEPF
Two-stage	0.597	0.304	-0.732	-0.639	-0.536	0.291
Prototype	0.431	0.420	-0.630	-0.340	-0.216	0.292

Table 15: Spearman rank correlation between SES indices and six variables of interest.

In four of the six cases, the two-stage index correlates much higher with these variables than does the prototype index. Two of the first three variables (that are commonly used in the development of SES indices) validate better with the two-stage approach. Hence we can conclude that the two-stage index validates better in most cases, relative to the prototype index, when considering these six proxy variables of interest. Similar to the results found with the complementary index, this is a good indication that our SES index is scoring the DAs in an intuitive manner and which can be easily interpreted.

5.4.2 Validation over the INSPQ

We next validated our method against the INSPQ and found that our SES DA scores correlated better with the first two components of the INSPQ relative to those with the prototype index, as shown in Table 16.

Index	INSPQ, material	INSPQ, social
Two-stage	0.589	0.494
Prototype	0.537	0.425

Table 16: Spearman rank correlation between SES indices and the first two components of the INSPQ.

Validation is better in both cases. This is a good indication that the two-stage index simultaneously captures more of the material-based behavior and social-based behavior of DAs than does the prototype index, and that utilizing only the first principal component(s) in each stage of the PCA should suffice in constructing an SES index.

5.4.3 Validation over the SEFI

Finally, we validated the SES dissemination scores from the two-stage index via the Spearman correlation with those scores based on the SEFI-like index, for a value of 0.403. The correlation of scores based on the prototype index and our SEFI-like index was 0.404. Validation with respect to the SEFI has done approximately equally as well with the two-stage index relative to the prototype index.

5.4.4 Further Validation

The Spearman rank correlation of the SES scores of the DAs based on our two-stage index and the prototype index scored at 0.67, indicating that there is close agreement in the scores between the two approaches. The validation scores for the two-stage index came out either as approximately equal to or better than those with the prototype index. We anticipate that validation of the two-stage index with the ORSEI and VANDIX indices should do well.

5.5 The Bootstrap Analysis

The bootstrap analysis (Efron and Tibshirani, 1993) is a resampling procedure that can be used to estimate the variance of estimators based on an observed sample. In our study, we have applied the bootstrap method to obtain confidence intervals for the final SES scores for the LHAs. Our method is as follows:

For each LHA L_i, where $i = 1,2,...,L$ and L is the number of LHAs, suppose that the number of DAs that are associated with L_i is n_i. Choose B, the number of bootstrap resamples, where B is sufficiently large.

Step 1: For each $i = 1,2,...,L$, sample n_i DAs *with replacement* from L_i.

Step 2: Proceed with the two-stage analysis as if these sampled DAs comprise the entirety of BC.

Step 3: Record the average scores from Step 2 for each LHA L_i.

Step 4: Repeat steps 1, 2, and 3 a total of B times.

Step 5: Draw inference on the distribution of the scores for the LHAs.

In our analysis we took $B = 5000$. Figure 8 provides the average SES scores by LHA in ascending order with the 95% bootstrap confidence intervals represented by the horizontal lines. A table of the scores is provided in section A.4 of the Appendix.

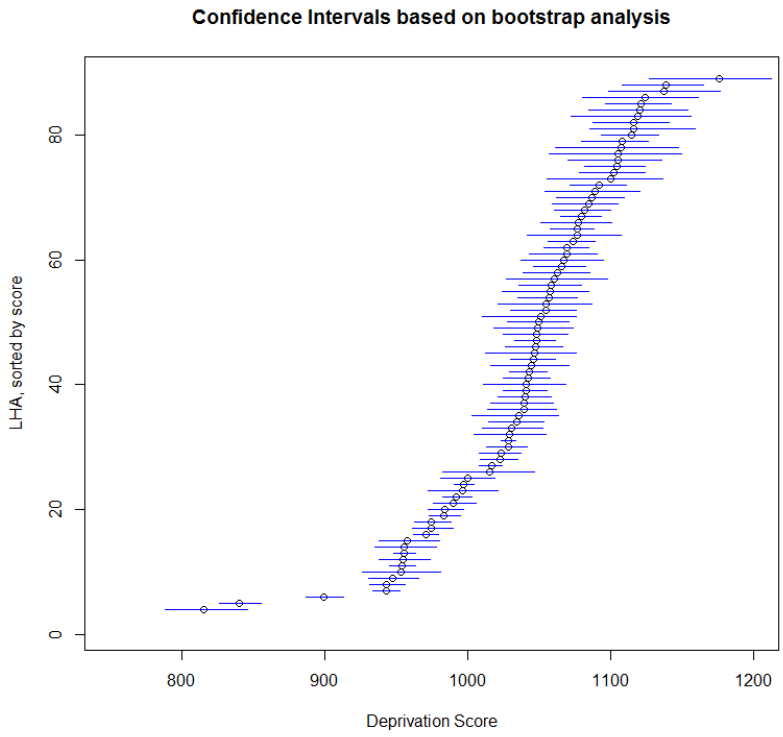

Figure 8: The bootstrap confidence intervals for the LHAs.

Figure 8 reveals some of the underlying variability in the average SES scores for the LHAs, indicating that there is some degree of uncertainty in the final average scores/rankings of the LHAs. Some of this can be attributed to the number of DAs (and therefore population density) affiliated with the LHAs, as shown in Figure 9. This figure gives a plot of the deviation of the bootstrapped scores against the number of DAs affiliated with the LHAs. Notice the smaller deviation in SES scores for LHAs that are affiliated with a larger number of DAs. The average standard deviation of the bootstrapped scores was 17.3.

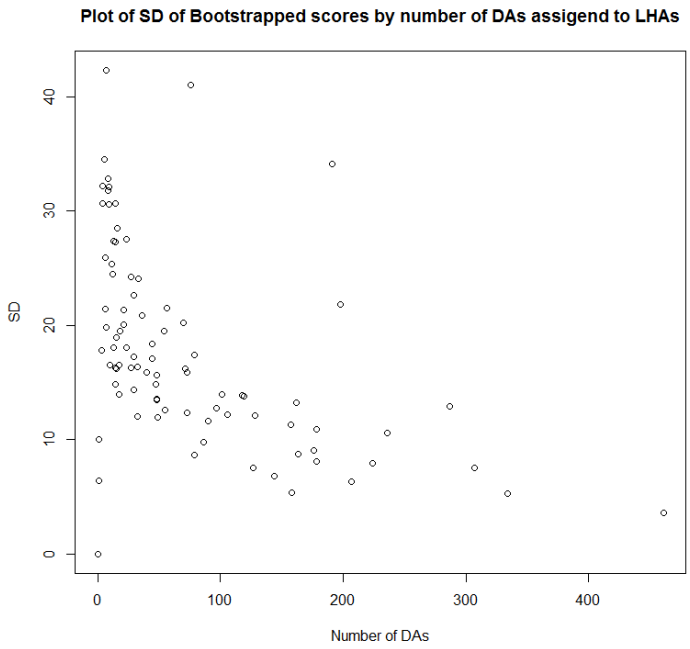

Figure 9: A plot of the deviation of the bootstrap scores for each LHA against the number of DAs affiliated with the LHA.

One suggested approach to utilizing the bootstrap scores is to choose a subset of the LHAs that are of interest, and while holding all other scores fixed, consider the possible range of scores for these LHAs. For example, to determine health authorities that are most deprived, consider the range of the confidence intervals for those LHAs where this range will exceed perhaps the fifth highest score, as shown in Figure 10.

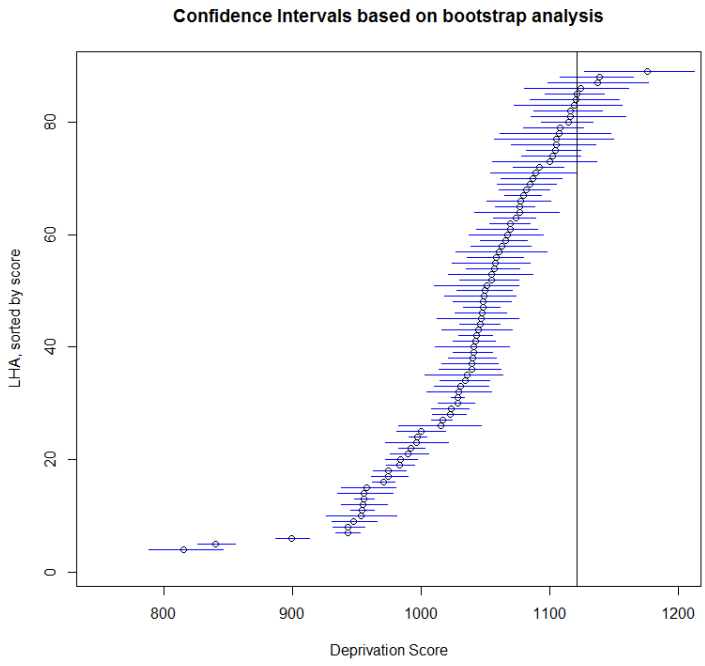

Figure 10: The bootstrap confidence intervals with the fifth highest ranked score indicated by the vertical line.

Given the approach taken in our bootstrap analysis, the variability found in the rankings can also reflect potential inconsistencies in the final scores between different SES indices. The bootstrap analysis reveals that some of the LHAs possess a high degree of variability in the average scores based on the two-stage index, and scores based on other measures of SES may prove to be as variable. Therefore, as there is no current gold standard of validation of SES indices, validation against SES indices found in the current literature should be taken with a grain of salt. It is strongly advised that future SES indices that are developed for use by the BC MoH contain an analysis that provides a measure of variability in the final scores/rankings.

6 Discussion

In this study we have proposed two socioeconomic indices to compare the SES of DAs and LHAs within BC. Our approaches to developing SES indices are different from some of those found in the existing literature (for example, the SEFI (1996), INSPQ (2008), and NZDep (Salmond and Crampton, 2002)) as well as those that have been applied to investigating the deprivation of areas of BC (ORSEI (BC Stats, 2011) and Vandix (Bell et al. (2008))). The primary difference is that, in contrast to the existing approaches, our SES indices do not rely on any subjective decisions as to which variables should be included in the analysis. In both cases, we start with the full data set and let our principal components analysis determine the amount of influence each variable has towards the final scores. Any inherent bias with a subjective approach is avoided with our methods.

There are some limitations in our study as census data from some DAs are not available, as noted in Section 2. The illustrations that show the trends in the missing data patterns reveal that there is a large quantity of census data missing from First Nations reserves. Therefore, we shall note that there is to be some expected bias in the final results of our study as it has been mentioned that such areas are more likely to be deprived (Human Early Learning Partnership, 2009) and hence the reason why some of the northern LHAs score well in our study.

There are several methodological problems with the existing prototype SES index that the MoH recently developed (and can be readily addressed using the methods outlined in this report). The result of the contrast between methods is known as the complementary SES index, and statistics suggest these methods are an improvement over existing prototype methods. The validation of the complementary SES index shows positive findings.

The two-stage methods may provide a viable alternative to deriving an SES index. This approach ensures that education variables are not over-represented in the index. Applying the two-stage approach also provides indications of the variability of the SES index, a concept not apparent in the prototype or complementary SES indices, primarily due to the computational expense required to evaluate the SES scores. Future work on using a bootstrap style of analysis to estimate the variability of the LHA rankings under the use of the complementary index would be valuable.

Recommendations

We recommend that the existing prototype SES index should be dropped by the BC MoH since its derivation has some methodological problems that cannot be addressed. As the two-stage index has done well in terms of validation and ranking LHAs, in conjunction with providing confidence intervals for the SES scores of the LHAs, we suggest that the BC MoH adopt the two-stage SES index described above and that the results based on this analysis should be used for policy and planning decisions regarding the provision of health services. We shall note that the BC MoH

should recognize that SES indices have underlying variability only reflected in the two-stage approach proposed in this report.

Notwithstanding these recommendations, the validation of either SES indices is problematic, and we cannot assume that our methods are representative of variation of SES among DAs in BC. Additional validation should proceed using independent data not available to this study team.

7 Acknowledgements

We would like to thank Jeet Rana of the BC MoH, who provided valuable input to our study. We would also like to thank Nadya Repin, Dawn Mooney, Trafford Crump, Rachael McKendry, and Scally Chu of the UBC Centre for Health Services and Policy Research for their valuable suggestions.

8 References

Aboriginal Peoples in Canada in 2006: Inuit, Métis and First Nations, 2006 Census. Available at: http://www12.statcan.ca/census-recensement/2006/as-sa/97-558/pdf/97-558-XIE2006001.pdf

BC Stats (2011) “British Columbia Regional Socio-Economic Indicators-Methodology" Victoria, BC

Bell, N., Schuurman, N., Oliver, L., and Hayes, M. (2007). Towards the construction of place-specific measures of deprivation: a case study from the Vancouver metropolitan area. *The Canadian Geographer*, Volume 51, number 4 (444-461).

Census of Canada. Cartographic Boundary Files, 2006. Available at: http://abacus.library.ubc.ca/jspui/handle/10573/41796

Efron, B., and Tibshirani, R. (1993). An Introduction to the Bootstrap. Boca Raton, FL: Chapman and Hall/CRC.

Human Early Learning Partnership (2009). SES Mapping Package. Vancouver: Human Early Learning Partnership. Available at: http://earlylearning.ubc.ca/media/uploads/mapsets/ses_t1t2/BC_SES/BC_SES_T1T2.pdf

INSPQ (2008) Statistics-Institut national de sante publique du Quebec (INSPQ) Available at: http://www.inspq.qc.ca/santescope/indicedefavoeng.asp?NoIndD=9&Lg=en

Jackson, J. (1991). A user’s guide to principal components. Los Angeles, John Wiley and Sons, Inc.

Joliffe, I. (2002). Principal component analysis. Second edition. New York, NY:Springer-Verlag.

McCabe, G. (1984). Principal Variables. *Technometrics*, Volume 26, Issue 2.

Messer, L., Laraia, B., Kaufman, J., Eyster, J., Holzman, C., Culhane, J., Elo, I., Burke, J., and O’Campo, P. (2006). The development of a standardized neighborhood deprivation index. *Journal of Urban Health*, Volume 83, number 6 (1041-1062).

Pampalon, R., and Raymon, G. (2000). A deprivation index for health and welfare planning in Quebec. *Chronic Diseases in Canada*, Volume 21, number 3 (104-113).

Regional Measures of Socioeconomic Status for the Use in the Provision of Health Services. (2012). Victoria, BC: British Columbia Ministry of Health.

Salmond, C., and Crampton, P. (2002). “NZDep2001 index of deprivation." Wellington: Wellington School of Medicine and Health Sciences

SEFI (1996) “Socioeconomic factor index" Available at: http://mchp-appserv.cpe.umanitoba.ca/viewDefinition.php?definitionID=103587

Appendix

A.1 Enlarged Portions of the Missing Data Map

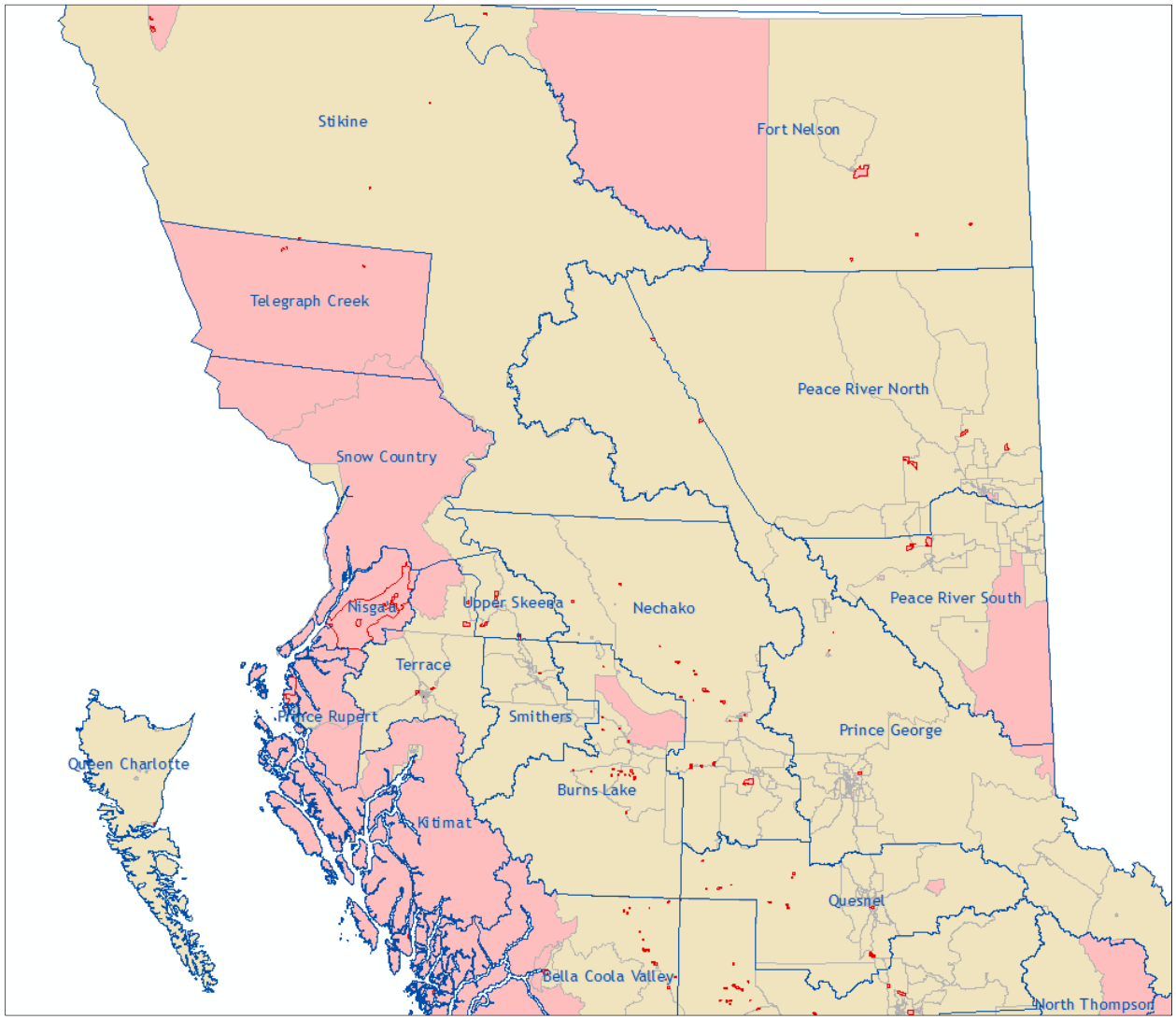

Figure A.1.1: DAs in Northern BC. Areas highlighted in brown are DAs that are included in our data set and the areas in pink are DAs that were not included in our data set. The areas outlined in red represent First Nations reserves.The blue outlines and names indicate LHAs.
Source: Statistics Canada.

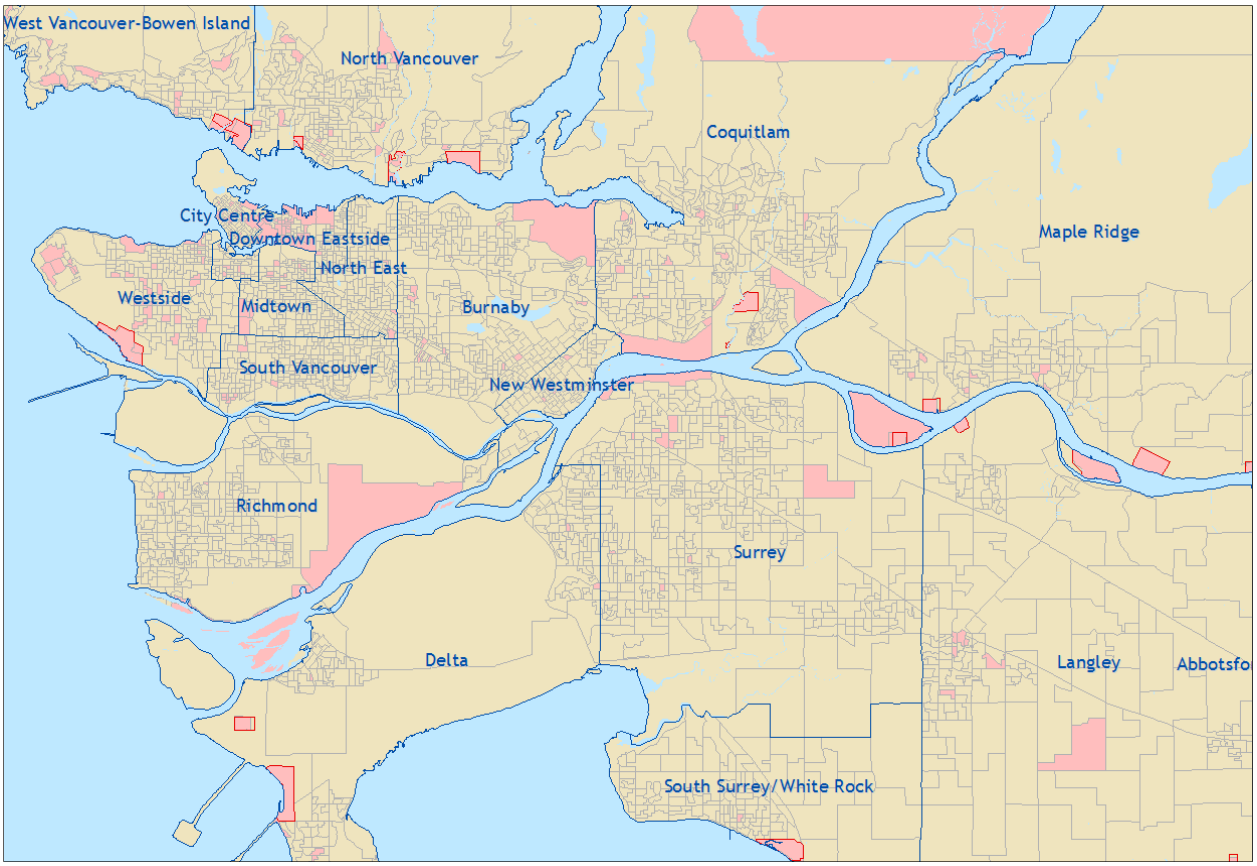

Figure A.1.2: DAs in Metro Vancouver. Areas highlighted in brown are DAs that are included in our data set and the areas in pink are DAs that were not included in our data set. The areas outlined in red represent First Nations reserves.The blue outlines and names indicate LHAs.
Source: Statistics Canada.

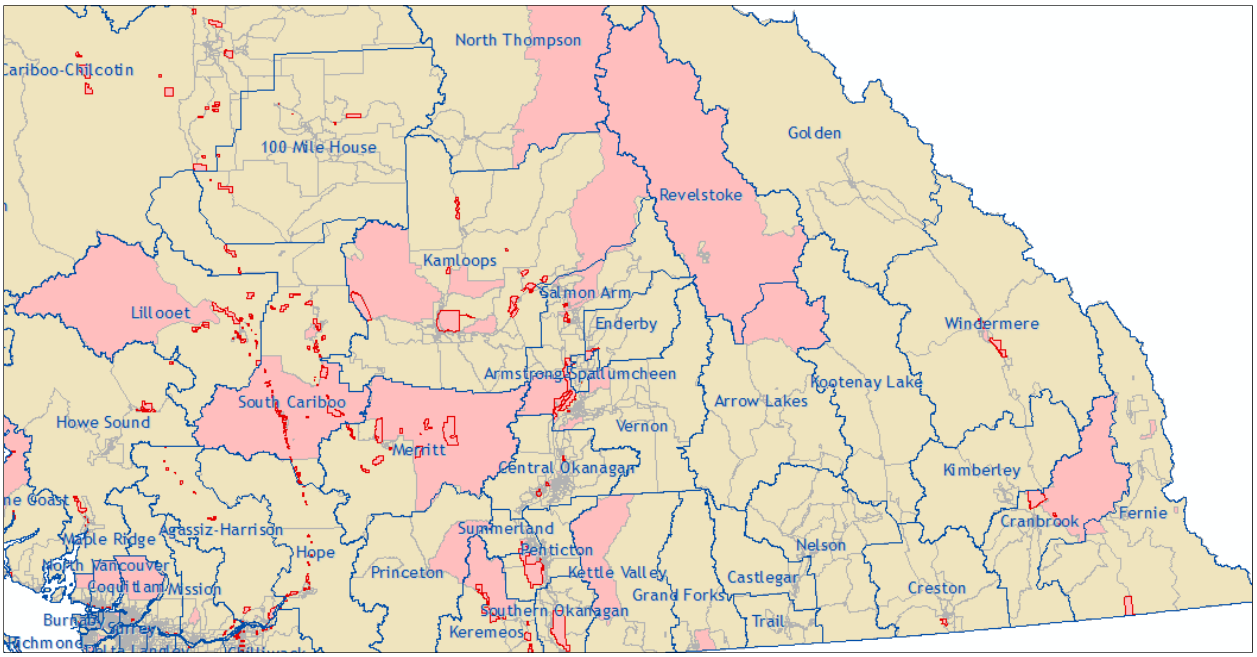

Figure A.1.3: DAs in the Interior of BC. Areas highlighted in brown are DAs that are included in our data set and the areas in pink are DAs that were not included in our data set. The areas outlined in red represent First Nations reserves.The blue outlines and names indicate LHAs.
Source: Statistics Canada.

LHA	True # DAs	Our # DAs	Proportion DAs	True Pop	Our Pop	Proportion Pop
100 Mile House	35	27	0.771	14165	13802	0.974
Abbotsford	186	176	0.946	124515	120899	0.971
Agassiz-Harrison	28	13	0.464	8608	7227	0.840
Alberni	77	56	0.727	30664	28200	0.920
Armstrong-Spallumcheen	19	18	0.947	9206	9201	0.999
Arrow Lakes	11	9	0.818	4470	4285	0.959
Bella Coola Valley	24	3	0.125	2950	1359	0.461
Burnaby	322	307	0.953	203331	193993	0.954
Burns Lake	43	12	0.279	7427	5888	0.793
Campbell River	91	71	0.780	39571	36590	0.925
Cariboo-Chilcotin	93	47	0.505	25694	22525	0.877
Castlegar	25	23	0.920	12466	12352	0.991
Central Coast	3	0	0.000	1348	0	0.000
Central Okanagan	243	224	0.922	162276	150676	0.929
Chilliwack	150	128	0.853	77102	72537	0.941
City Centre	144	119	0.826	104258	89209	0.856
Coquitlam	306	287	0.938	196845	192041	0.976
Courtenay	114	106	0.930	59482	56974	0.958
Cowichan	107	90	0.841	54007	50626	0.937
Cranbrook	50	44	0.880	23771	22835	0.961
Creston	25	23	0.920	12450	12152	0.976
Delta	169	162	0.959	97402	94862	0.974
Downtown Eastside	93	73	0.785	53622	38769	0.723
Enderby	14	13	0.929	7272	6919	0.951
Fernie	34	29	0.853	14583	14064	0.964
Fort Nelson	17	10	0.588	6260	5593	0.893
Golden	14	14	1.000	6908	6908	1.000
Grand Forks	15	14	0.933	8647	7943	0.919
Greater Victoria	364	334	0.918	210921	196596	0.932
Gulf Islands	19	17	0.895	14741	14741	1.000
Hope	39	14	0.359	8089	7454	0.921
Howe Sound	69	48	0.696	30975	27161	0.877
Kamloops	182	157	0.863	102562	94975	0.926
Keremeos	14	7	0.500	4908	3878	0.790
Kettle Valley	6	5	0.833	3480	2917	0.838
Kimberley	18	17	0.944	7870	7870	1.000
Kitimat	19	14	0.737	10080	8762	0.869
Kootenay Lake	6	6	1.000	2997	2997	1.000
Ladysmith	37	29	0.784	16840	15668	0.930

Lake Cowichan	16	11	0.688	6082	5868	0.965
Langley	188	178	0.947	117858	113753	0.965
Lillooet	30	6	0.200	4250	2899	0.682
Maple Ridge	136	127	0.934	85560	83554	0.977
Merritt	28	16	0.571	10951	8777	0.801
Midtown	156	144	0.923	83909	79195	0.944
Mission	79	73	0.924	37897	36193	0.955
Nanaimo	188	178	0.947	95880	93178	0.972
Nechako	58	27	0.466	15363	12690	0.826
Nelson	49	48	0.980	23500	23500	1.000
New Westminster	90	86	0.956	58549	56501	0.965
Nisga'a	7	0	0.000	1919	0	0.000
North East	162	158	0.975	95396	91110	0.955
North Thompson	10	8	0.800	4166	3823	0.918
North Vancouver	216	198	0.917	129776	121455	0.936
Peace River North	63	55	0.873	32099	29589	0.922
Peace River South	57	48	0.842	25926	24569	0.948
Penticton	75	70	0.933	39461	37066	0.939
Powell River	40	36	0.900	19240	18082	0.940
Prince George	175	163	0.931	92503	90057	0.974
Prince Rupert	34	29	0.853	14278	13392	0.938
Princeton	11	8	0.727	4885	4372	0.895
Qualicum	82	79	0.963	43110	42289	0.981
Queen Charlotte	11	7	0.636	4812	3332	0.692
Quesnel	64	44	0.688	22449	21608	0.963
Revelstoke	16	15	0.938	7936	7230	0.911
Richmond	241	236	0.979	174461	172922	0.991
Saanich	104	97	0.933	60525	56886	0.940
Salmon Arm	65	54	0.831	32214	28952	0.899
Smithers	35	32	0.914	15453	15263	0.988
Snow Country	3	1	0.333	542	496	0.915
Sooke	92	79	0.859	58977	56354	0.956
South Cariboo	91	9	0.099	6823	4241	0.622
South Surrey/White Rock	121	118	0.975	77616	77014	0.992
South Vancouver	218	207	0.950	127405	120481	0.946
Southern Okanagan	36	33	0.917	18671	17002	0.911
Stikine	7	1	0.143	996	439	0.441
Summerland	24	21	0.875	11550	10465	0.906
Sunshine Coast	88	49	0.557	27759	26297	0.947
Surrey	475	461	0.971	336369	331193	0.985
Telegraph Creek	6	0	0.000	677	0	0.000
Terrace	48	40	0.833	19837	18168	0.916

Trail	32	32	1.000	18615	18615	1.000
Upper Skeena	21	4	0.190	5236	1831	0.350
Vancouver Island North	45	21	0.467	11919	9318	0.782
Vancouver Island West	14	4	0.286	2274	1728	0.760
Vernon	112	101	0.902	60823	55516	0.913
West Vancouver-Bowen Island	89	76	0.854	49313	43027	0.873
Westside	220	191	0.868	125653	109445	0.871
Windermere	18	15	0.833	9261	8119	0.877

Table A.1.1: DA and Population Counts included in the master file and our data set, by LHA

A.2 Mathematical Details of McCabe's (1984) Principal Variables Criteria

With the principal variable approach, McCabe (1984) mentions that a specific subset of variables should be retained from the full set of variables when they optimize a choice of one of the following four functions of interest of the eigenvalues (variances of the principal components) associated with the conditional correlation matrix of the variables:

1) Minimize $\prod_{j=1}^{m^*}\theta_j$

2) Minimize $\sum_{j=1}^{m^*}\theta_j$

3) Minimize $\sum_{j=1}^{m^*}\theta_j^2$

4) Minimize $\sum_{j=1}^{m^-}\rho_j^2$

where $\theta_j, j = 1,2,...,m$ are the eigenvalues of the conditional correlation matrix of the m^* deleted variables, given the values of the m selected variables, and $\rho_j, j = 1,2,...,m^- = \min(m, m^*)$ are the canonical correlations between the set of m^* deleted variables and the set of m selected variables. McCabe mentions that such a subset of the original variables is termed a set of *principal variables*. As $m = 12$ and 43 variables are utilized at this stage of the analysis, there were 43 choose 12 different combinations that needed to be evaluated for this approach. Therefore we decided on the criteria (that is, criteria 2) for which McCabe (1984) has developed a step-wise variable selection procedure that chooses a subset of variables to retain where these variables should be close to meeting this criteria.

A.3 Final scores for the LHAs based on the Complementary Index

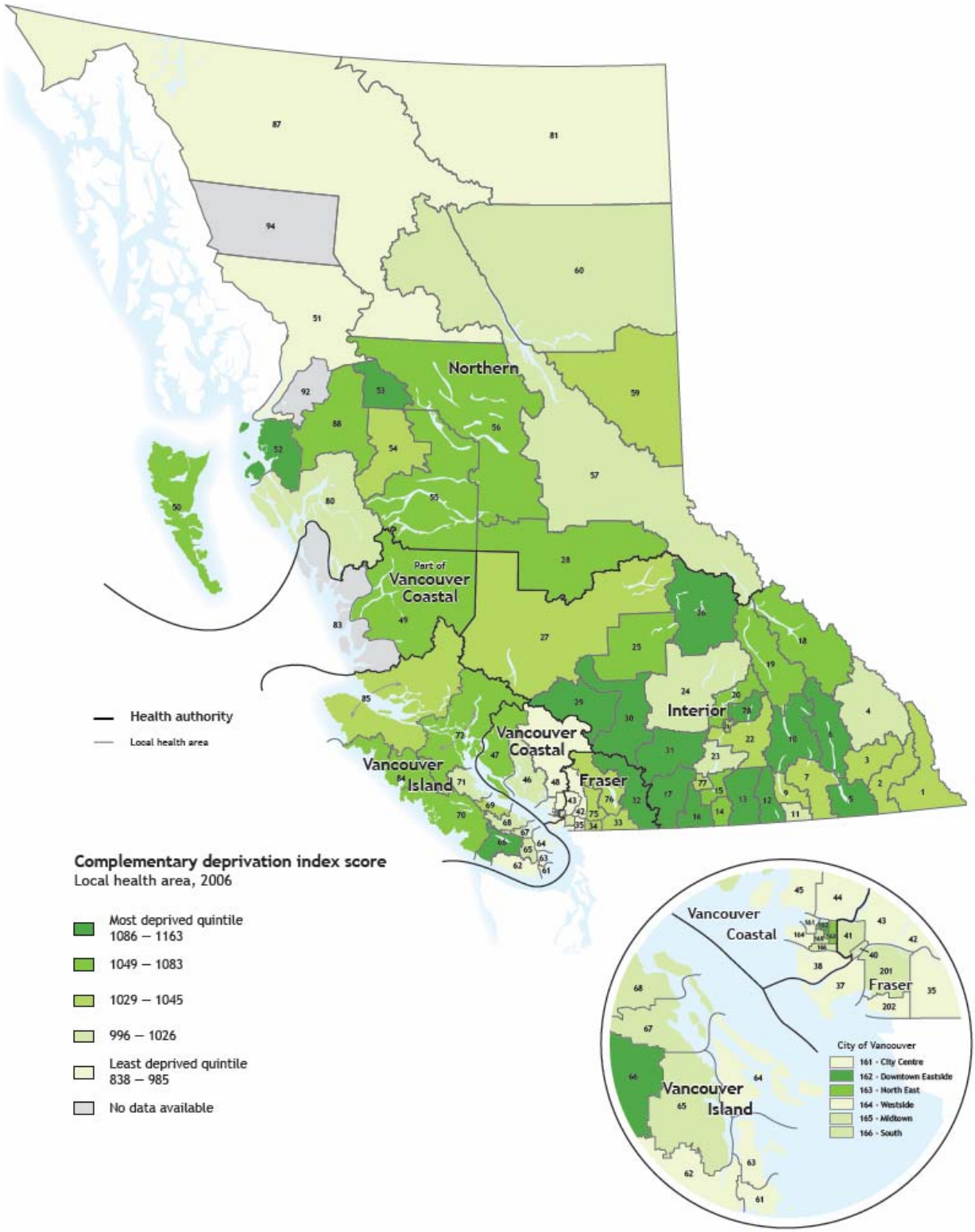

Figure A.3.1: Illustration of complementary index deprivation scores of LHAs categorized by quintiles.

Rank	Score	Local Health Area
NN	NaN	Central Coast
NaN	NaN	Nisga'a
NaN	NaN	Telegraph Creek
1	837.512	West Vancouver-Bowen Island
2	852.929	Westside
3	901.680	North Vancouver
4	917.060	Saanich
5	919.765	Stikine
6	934.203	South Surrey/White Rock
7	939.041	City Centre
8	947.831	Delta
9	949.404	Fort Nelson
10	952.521	Coquitlam
11	958.620	Howe Sound
12	958.997	Sooke
13	966.070	Greater Victoria
14	971.604	Langley
15	972.648	Snow Country
16	975.667	Gulf Islands
17	980.551	Richmond
18	984.692	Maple Ridge
19	996.093	Burnaby
20	996.286	Windermere
21	998.094	New Westminster
22	1004.568	Sunshine Coast
23	1008.597	Peace River North
24	1008.928	Cowichan
25	1010.649	Midtown
26	1010.888	Central Okanagan
27	1015.915	Courtenay
28	1017.355	Trail
29	1022.903	Nanaimo
30	1023.228	South Vancouver
31	1023.477	Prince George
32	1025.847	Ladysmith
33	1026.046	Surrey
34	1026.156	Kamloops
35	1026.165	Kitimat
36	1029.253	Vernon
37	1029.652	Summerland
38	1030.009	Abbotsford
39	1031.312	Fernie
40	1032.287	Smithers
41	1033.804	Cranbrook
42	1034.047	Qualicum
43	1034.301	Chilliwack

44	1034.344	Nelson
45	1034.882	Vancouver Island North
46	1035.036	Peace River South
47	1037.855	Mission
48	1041.427	Castlegar
49	1043.902	Agassiz-Harrison
50	1044.662	Armstrong-Spallumcheen
51	1045.070	Kimberley
52	1045.321	Cariboo-Chilcotin
53	1048.594	Queen Charlotte
54	1050.004	Penticton
55	1052.500	Terrace
56	1052.550	Vancouver Island West
57	1053.678	Golden
58	1056.564	Salmon Arm
59	1058.513	Nechako
60	1059.205	Campbell River
61	1062.404	North East
62	1062.685	Powell River
63	1066.545	Southern Okanagan
64	1067.915	Bella Coola Valley
65	1072.005	Revelstoke
66	1072.037	Quesnel
67	1074.774	100 Mile House
68	1077.631	Alberni
69	1082.545	Burns Lake
70	1085.963	Arrow Lakes
71	1086.608	Lillooet
72	1086.654	Grand Forks
73	1087.720	Downtown Eastside
74	1088.558	Enderby
75	1088.900	Kettle Valley
76	1091.122	Prince Rupert
77	1092.015	Hope
78	1093.789	Kootenay Lake
79	1102.175	Merritt
80	1102.553	Creston
81	1111.362	South Cariboo
82	1114.876	Lake Cowichan
83	1116.578	North Thompson
84	1117.517	Princeton
85	1130.890	Keremeos
86	1163.425	Upper Skeena

Table A.3.1: Final LHA Scores by Complementary Index

A.4 Final scores for the LHAs based on the Two-Stage Principal Components Analysis Index

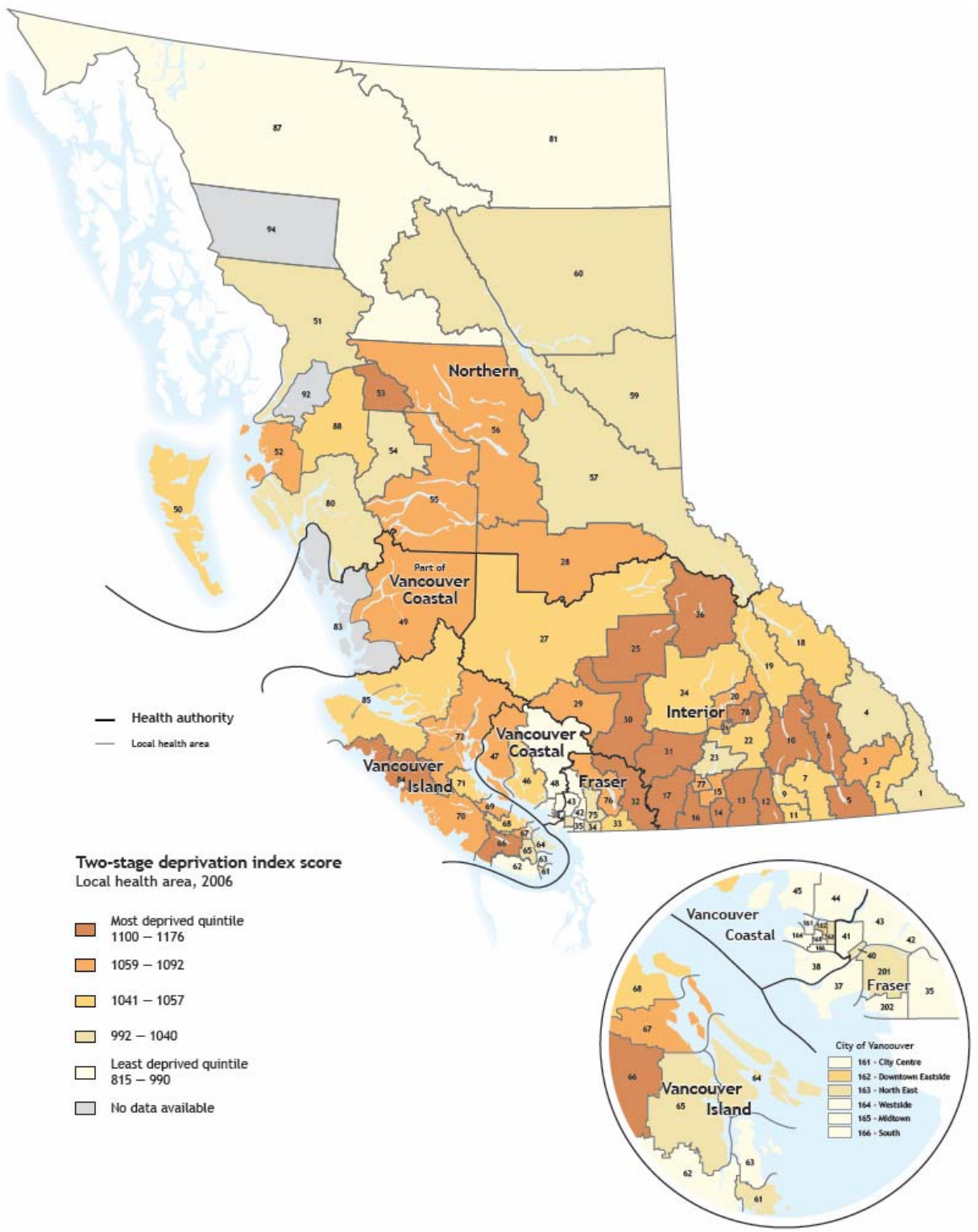

Figure A.4.1 Illustration of two-stage index deprivation scores of LHAs categorized by quintiles.

Rank	Score	Local Health Area
NaN	NaN	Central Coast
NaN	NaN	Nisga'a
NaN	NaN	Telegraph Creek
1	814.877	West Vancouver-Bowen Island
2	840.010	Westside
3	899.161	North Vancouver
4	942.766	Coquitlam
5	943.315	Delta
6	947.443	City Centre
7	953.563	Fort Nelson
8	953.773	Richmond
9	954.909	Saanich
10	955.217	Stikine
11	955.278	South Surrey/White Rock
12	957.489	Howe Sound
13	970.502	Burnaby
14	974.616	Sooke
15	974.771	Langley
16	983.376	South Vancouver
17	984.330	Midtown
18	990.137	Maple Ridge
19	992.232	Greater Victoria
20	996.633	Peace River North
21	997.257	Surrey
22	999.919	New Westminster
23	1015.073	Windermere
24	1016.780	North East
25	1022.564	Central Okanagan
26	1023.784	Prince George
27	1028.361	Abbotsford
28	1028.736	Snow Country
29	1029.612	Gulf Islands
30	1030.857	Smithers
31	1034.516	Cowichan
32	1036.013	Kitimat
33	1039.188	Fernie
34	1039.196	Peace River South
35	1040.408	Mission
36	1040.661	Kamloops
37	1041.157	Trail
38	1042.221	Sunshine Coast

39	**1043.154**	**Nanaimo**
40	**1044.610**	**Terrace**
41	**1046.256**	**Chilliwack**
42	**1046.867**	**Castlegar**
43	**1047.519**	**Vernon**
44	**1047.948**	**Courtenay**
45	**1048.631**	**Cariboo-Chilcotin**
46	**1048.972**	**Cranbrook**
47	**1049.783**	**Golden**
48	**1050.927**	**Queen Charlotte**
49	**1054.544**	**Downtown Eastside**
50	**1055.014**	**Vancouver Island North**
51	**1057.092**	**Nelson**
52	**1057.479**	**Revelstoke**
53	**1058.734**	**Nechako**
54	**1061.092**	**Summerland**
55	**1062.999**	**Ladysmith**
56	**1065.619**	**Campbell River**
57	**1067.249**	**Armstrong-Spallumcheen**
58	**1069.242**	**Quesnel**
59	**1069.575**	**Kimberley**
60	**1074.210**	**Qualicum**
61	**1076.796**	**Prince Rupert**
62	**1076.847**	**Bella Coola Valley**
63	**1077.322**	**Penticton**
64	**1079.731**	**Agassiz-Harrison**
65	**1081.978**	**Salmon Arm**
66	**1084.882**	**Powell River**
67	**1087.144**	**Lillooet**
68	**1089.415**	**Burns Lake**
69	**1091.969**	**Alberni**
70	**1099.972**	**Merritt**
71	**1102.479**	**Southern Okanagan**
72	**1104.636**	**100 Mile House**
73	**1105.117**	**Enderby**
74	**1105.142**	**Vancouver Island West**
75	**1107.319**	**Hope**
76	**1108.370**	**Lake Cowichan**
77	**1114.804**	**Kootenay Lake**
78	**1116.095**	**Upper Skeena**
79	**1116.390**	**Grand Forks**
80	**1119.525**	**North Thompson**

81	**1120.813**	**Arrow Lakes**
82	**1121.382**	**Creston**
83	**1124.393**	**Princeton**
84	**1137.804**	**Kettle Valley**
85	**1139.181**	**South Cariboo**
86	**1176.309**	**Keremeos**

Table A.4.1: Final LHA Scores by Two-stage Index

Two-Stage Index

LHA	Lower Bound	Upper Bound
Central Coast	NaN	NaN
Nisga'a	NaN	NaN
Telegraph Creek	NaN	NaN
West Vancouver-Bowen Island	788.366	845.644
Westside	826.299	855.128
North Vancouver	886.610	913.344
Coquitlam	933.818	952.897
Delta	931.522	956.230
City Centre	930.941	965.973
Fort Nelson	925.917	981.195
Richmond	945.111	963.701
Saanich	938.152	973.798
Stikine	948.450	963.390
South Surrey/White Rock	935.335	978.344
Howe Sound	937.652	980.217
Burnaby	962.166	979.704
Sooke	961.634	989.965
Langley	962.939	988.219
South Vancouver	973.137	995.108
Midtown	972.618	997.156
Maple Ridge	976.077	1006.098
Greater Victoria	982.520	1002.968
Peace River North	972.014	1021.593
Surrey	990.357	1004.385
New Westminster	980.911	1019.295
Windermere	982.376	1046.661
North East	1008.081	1024.377
Central Okanagan	1008.761	1034.834
Prince George	1008.030	1037.670
Abbotsford	1013.314	1041.878
Snow Country	1023.656	1033.374
Gulf Islands	1004.488	1054.952
Smithers	1010.240	1052.354
Cowichan	1014.717	1053.659
Kitimat	1003.075	1063.502
Fernie	1013.743	1062.352
Peace River South	1016.098	1060.068
Mission	1021.239	1058.853
Kamloops	1024.957	1055.288

Trail	1011.123	1068.616
Sunshine Coast	1024.870	1058.107
Nanaimo	1029.346	1055.278
Terrace	1016.188	1070.758
Chilliwack	1029.990	1061.228
Castlegar	1012.429	1075.847
Vernon	1026.282	1066.741
Courtenay	1033.182	1061.543
Cariboo-Chilcotin	1024.650	1070.043
Cranbrook	1018.370	1074.199
Golden	1027.836	1071.187
Queen Charlotte	1010.350	1075.931
Downtown Eastside	1030.067	1075.979
Vancouver Island North	1020.900	1087.114
Nelson	1034.939	1076.529
Revelstoke	1024.191	1085.086
Nechako	1035.770	1079.911
Summerland	1027.055	1098.169
Ladysmith	1038.761	1085.477
Campbell River	1046.305	1082.630
Armstrong-Spallumcheen	1037.522	1094.931
Quesnel	1042.959	1090.593
Kimberley	1053.329	1084.981
Qualicum	1056.672	1089.203
Prince Rupert	1041.624	1107.830
Bella Coola Valley	1057.992	1088.731
Penticton	1051.219	1100.927
Agassiz-Harrison	1064.846	1093.509
Salmon Arm	1060.825	1100.495
Powell River	1059.360	1105.558
Lillooet	1061.977	1109.342
Burns Lake	1054.472	1120.672
Alberni	1071.385	1111.141
Merritt	1055.454	1136.637
Southern Okanagan	1077.967	1123.981
100 Mile House	1081.558	1124.686
Enderby	1070.230	1135.786
Vancouver Island West	1056.952	1149.950
Hope	1061.699	1147.635
Lake Cowichan	1079.964	1126.270
Kootenay Lake	1093.650	1133.980
Upper Skeena	1085.371	1159.400

Grand Forks	1087.900	1141.087
North Thompson	1072.325	1156.552
Arrow Lakes	1085.109	1154.365
Creston	1096.355	1142.252
Princeton	1080.174	1161.924
Kettle Valley	1099.044	1176.891
South Cariboo	1108.434	1165.605
Keremeos	1127.589	1212.566

Table A.4.2: Bootstrap Confidence Interval Scores from the Bootstrap Analysis of the Two Stage Index.