Applied Research Branch
Strategic Policy
Human Resources Development Canada

Direction générale de la recherche appliquée
Politique stratégique
Développement des ressources humaines Canada

New COPS Occupational Projection Methodology

T-99-3E

by

Richard Archambault

October 1999

■

The Technical Paper Series includes studies and research conducted under the auspices of the Applied Research Branch of Strategic Policy. Papers published in this series incorporate research of a technical nature intended for a limited number of specialists in particular areas of human resources studies.

La série des documents techniques comprend des études et des travaux de recherche réalisés sous l'égide de la Direction générale de la recherche appliquée, Politique stratégique. Il s'agit notamment de recherches analytiques techniques destinées à un nombre restreint de spécialistes dans des domaines particuliers liés aux ressources humaines.

■

This document is a translation from French. Although the English version has been carefully prepared, the original document should be taken as correct./
Le présent document a été traduit du français. Bien que la version anglaise ait été préparée avec soin, le document original fait foi.

This paper is available in French under the title "Nouvelle méthode de projection des parts professionnelles du SPPC."/
La version française de ce document est disponible sous le titre « Nouvelle méthode de projection des parts professionnelles du SPPC ».

■

Publication Date/Date de parution 2000
ISBN: 0-662-28483-6
Cat. No./N° de cat. MP32-30/99-3E

■

General enquiries regarding the documents published by the Applied Research Branch should be addressed to:

Publications Office
Applied Research Branch
Strategic Policy
Human Resources Development Canada
165 Hôtel de Ville St., Phase II, 7th Floor
Hull, Quebec, Canada
K1A 0J2

Telephone: (819) 994-3304
Facsimile: (819) 953-9077
E-mail: research@spg.org
http://www.hrdc-drhc.gc.ca/arb/

Si vous avez des questions concernant les documents publiés par la Direction générale de la recherche appliquée, veuillez communiquer avec :

Service des publications
Direction générale de la recherche appliquée
Politique stratégique
Développement des ressources humaines Canada
165, rue Hôtel de Ville, Phase II, 7e étage
Hull (Québec) Canada
K1A 0J2

Téléphone : (819) 994-3304
Télécopieur : (819) 953-9077
Courrier électronique : research@spg.org
http://www.hrdc-drhc.gc.ca/dgra/

Abstract

The purpose of this paper is to evaluate the appropriateness and sensitivity of the methodological framework currently used for the Canadian Occupational Projection System (COPS) projections of occupational demand. More specifically, this paper analyses whether the current methodology used for COPS projections of occupational demand enables us to fully take into consideration the economic information (economic cycles and trends) contained in the occupational series, while minimizing the risk of introducing bias into the projections.

Examination of these issues has led to two recommendations: i) use a new projection method which takes the economic information contained in the occupational series into account to a greater degree, and ii) use a higher level of aggregation for the shares series by occupation and industry to minimize the risk of introducing statistical bias into the projections. The impact of the proposed changes is assessed on the basis of the performance of national and provincial projections.

The results of the evaluation indicate that the proposed new methodology produces better results in terms of projection performance than the results obtained with the current COPS methodology, at both the national and provincial levels. The recommended changes would increase the sensitivity of the share projections while providing the same level of detail in the occupational demand projections as does the current methodology.

Résumé

L'objectif de ce document est d'évaluer la pertinence et l'acuité du cadre méthodologique actuellement utilisé pour projeter la demande d'emploi par professions du Système de projection professionnelle canadienne (SPPC). Plus spécifiquement, ce document examine si la méthodologie actuellement utilisée pour projeter la demande d'emploi par professions du SPPC permet de bien prendre en considération l'information de nature économique (cycle et tendance), contenue dans les séries d'emplois par professions, tout en minimisant le risque d'introduire un biais dans la projection.

L'examen de ces questions conduit à deux recommandations: i) l'utilisation d'une nouvelle méthode de projection où l'information de nature économique contenue dans les séries d'emplois sera mieux prise en considération et ii) un niveau d'agrégation plus élevé des séries de parts par professions et industries qui permettra de minimiser le risque d'introduire un biais statistique dans la projection. L'impact des changements proposés est évalué sur la performance des projections nationales et provinciales.

Les résultats de l'évaluation indiquent que la nouvelle méthode proposée obtient de meilleurs résultats au niveau de la performance prévisionnelle que la méthodologie actuellement utilisée par le SPPC et ce autant au niveau national que provincial. En plus d'augmenter l'acuité des projections de parts, les recommandations proposées permettent de garder le même niveau de détail des projections d'emplois par professions que ceux de la méthodologie actuelle.

Acknowledgements

We wish to thank Christian Dea, Mario Lapointe and Richard Roy for their very helpful comments, and other staff members of the Applied Research Branch who made insightful remarks at the various seminars that led to the production of this document.

Table of Contents

1. Introduction

The mandate of the Canadian Occupational Projection System (COPS) is to provide the best possible information on current and future conditions of labour supply and demand by occupation and industry in order to heighten job market effectiveness. The purpose of this paper is to evaluate the appropriateness and sensitivity of the methodological framework presently used for COPS projections of occupational demand.

The quality of COPS projections essentially depends on two components: the industrial occupational projection from the macro industrial scenario and the COPS occupational share projection used to allocate employment by occupation within industries. This paper attempts to answer two questions raised by the current COPS methodology to forecast occupational share:

- Does the current methodology enable us to fully take into consideration the economic information contained in the occupational series?
- Is the current level of disaggregation, 67 industries by 139 occupations, appropriate in terms of minimizing the risk of introducing a bias into projections given the small size of the Labour Force Survey (53 000 households)?

This paper is organized as follows. Section 2 contains a brief review of the evolution of the methodology used for making COPS projections and the two issues associated with labour demand projections. Section 3 addresses the question of minimizing the risk of introducing statistical bias into the projections. The problems associated with the current method (smoothing historical share series) are discussed and an alternative methodology is offered. The second issue – how to take economic information into account – is discussed in Section 4. Various alternatives are offered and the predictive performance of these alternative methods at the national and provincial levels, in comparison with the current methodology, is evaluated. In Section 5, the national occupational demand projections obtained with the current methodology are compared with those from the alternative methodology selected by this analysis. Conclusions and recommendations are offered in Section 6.

2. COPS Labour Demand Projection: Historical Development and Issues

The development of the COPS labour demand projection model should, inasmuch as possible, take into account the variability of occupational shares over time and determine the part of this variability owing to economic factors (labour market dynamics, trends, cycles) and that owing to statistical factors (noise).

In the past, major limitations in the available data made this impossible. There have therefore been two major stages in the development of the model.

1980s: Fixed share projection model

In the 1980s, time series were not available and there were insufficient observations over time, making it impossible to use a variable share model. Only information from one or two censuses was available.[1] With only one or two observation points, it was difficult to extract the dynamic behavior of occupational shares. As a result, COPS calculated occupational share projections by industry on a fixed share model (with the shares based on their latest historic values).[2] In other words, the shares by occupation and industry were held constant for the entire projection period.

1990s: Variable share projection model introduced

In the early 90s, with several censuses available, it was possible to obtain more information on occupational share trends. COPS therefore made a significant change in the occupational share model by introducing a methodology which let occupational shares vary over the projection period. This change attempted to take into consideration the substitution or complementarity

[1] The 1971 census was the first to provide information on occupations by industry. That information was not collected in the 1976 census. Since 1981, the information has been available at five-year intervals.

[2] This approach has been evaluated and it was shown that keeping coefficients fixed could cause projection errors greater than 30% for certain occupations. See the COPS publication entitled AFixed Coefficients: Are They Accurate Predictors for Occupational Employment?

effect that can exist between certain occupations over time. The new method calculated an occupation's linear trend from the last two available censuses and extended this by linear extrapolation over the projection period.[3] This variable share methodology had to be used with caution because linear trends calculated between two census years could be biased for two reasons: censuses are not always taken at the same point in the economic cycle, and the high disaggregation level used affects the reliability of information taken from censuses.

COPS started modeling the probable influence of the economic cycle as part of the labour demand model in 1995. This became possible with the newly available occupational share time series calculated from the industrial and occupational series in the Labour Force Survey (LFS).[4] COPS developed a methodology for determining which occupations were sensitive to the economic cycle.[5] A simple equation incorporating a trend and a cyclical variable was used to project the occupations sensitive to the economic cycle.[6] Exponential smoothing was used for projecting non-cyclical occupations.

Since 1997, COPS has no longer been modeling the influence of the economic cycle and has been using smoothing techniques to project all occupational share series by industry. The practice of modeling the influence of the economic cycle was dropped because so few occupations (ten or so) were identified as cyclical. Occupational share projection by industry is now calculated by a choice of three equations: exponential smoothing, an equation with linear trend and an equation with quadratic trend.[7]

[3] See the COPS technical paper entitled "Occupational Projection: Variable Coefficient Method."
[4] The historical series begin in 1984 and end with the last available year, 1998 for the year 1999.
[5] See the COPS technical paper entitled Canadian Occupational Projection System: A New Model of Occupational Coefficients," T-95-6, July 1995.
[6] The cyclical variable is defined as the output gap, which is estimated from the industrial GDP series. The trend entering into the output gap calculation is estimated by an HP filter.
[7] This choice is based on the following criteria: the equation used has to minimize the sum of the squares of the residues and at least one of the other two equations must produce a projection where the growth rate is of the same sign. When this choice is the quadratic equation, an average is then calculated with this equation and the linear equation if the growth rate of the projection is of the same sign: otherwise, the average is calculated with the exponential smoothing equation.

Box 2.1 summarizes the stages in producing national occupational projections.

Box 2.1 COPS National Occupational Projections Production Stages

1 Raw data manipulation:

a) Extracting industrial occupational series from the LFS using the 3-digit National Occupational Classification (NOC).
b) Converting occupational series to occupational shares.

2 Deriving occupational share projections:

a) Estimating and projecting the 3-digit NOC shares by smoothing.
b) Adjusting to guarantee that all projections are positive.
c) Imposing a growth floor or ceiling to limit the probability of projections where growth (positive or negative) would be implausible (too strong).
d) Normalizing share time series in each industry to guarantee that the sum of shares is equal to 1.

3 Processing share projections into job projections:

a) Calculating industrial and occupational employment by multiplying the share projections for each industry by the employment projection in the macro industrial scenario.
b) Calculating the total occupational projection by adding the employment for each occupation in the 67 industries.

Note: It is important to note that, since 1999, it has no longer been necessary to convert occupations from the Standard Occupational Classification (SOC) to the National Occupational Classification (NOC) since occupational series are now obtained directly from Statistics Canada in NOC form.

3. First Issue: Minimize the Risk of Introducing Statistical Bias into the Projections

3.1 Current methodology: the smoothing of occupational shares series

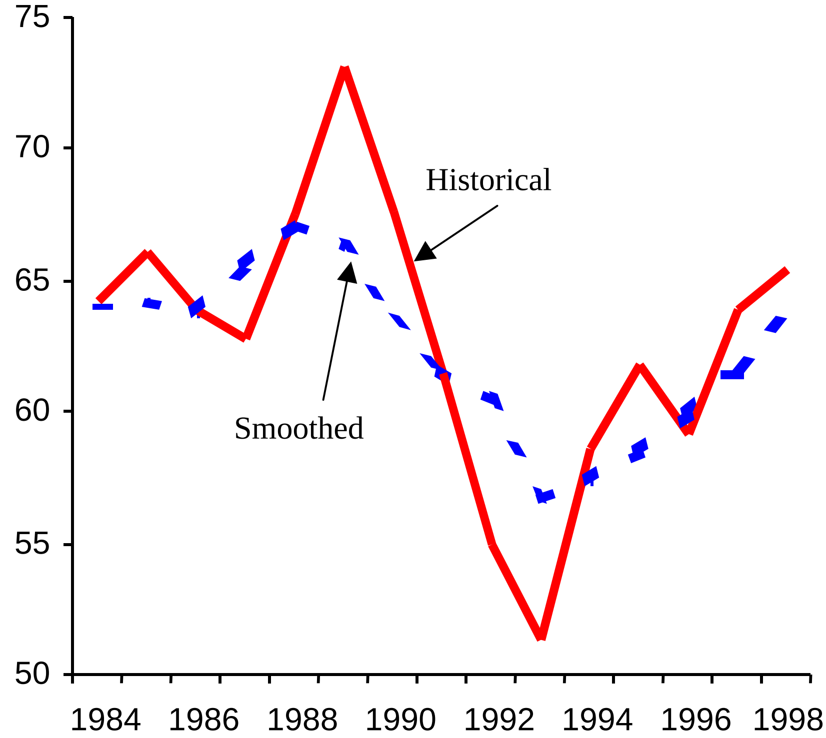
Nationally, the number of industrial occupational share series is equal to 9,313 or the product of 139 occupations by 67 industries.[8] Estimates of job numbers for each of these series come from the Labour Force Survey, a sample of 53,000 households or an average of slightly fewer than 6 households surveyed by observation. This high disaggregation level for a sample, the size of the LFS, produces a high level of statistical noise that may affect the reliability of information used. The introduction of smoothing by COPS in 1995 was intended to minimize the statistical noise in the occupational share series.

Smoothing reduces the volatility of occupational series. However, it does not enhance information quality concerning the trend and short-term behavior of the series. As well, smoothing may cause an arbitrary reallocation of occupational share within and throughout the 67 industries, which also changes the historical behavior of the share series. The result of this reallocation may be insignificant in one industry, but when aggregated, the sum of the errors generated by smoothing in each industry may cause serious projection errors.

Moreover, smoothing does not enable us to fully consider the economic information present in the share series. An important consequence of smoothing is the disappearance of certain economic shifts. In Figure 3.1, the smoothed series reaches its peak in 1988, one year earlier than the historical series, and produces a decline in the recession of the early 90s that is relatively 2 times smaller than what really happened in this occupation. As well, the smoothed series completely misses the slowdown in economic activity that occurred in 1996.

[8] Provincially, we have slightly over 34,000 series by province according to the 4-digit NOC, which includes 509 occupations.

Figure 3.1 Jobs (000) for machine operators in food processing, beverages … (946)



3.2 Alternative method: use a higher level of aggregation

There is an alternative solution to partly correct this situation. As previously stated, the problem stems from the smoothing introduced to reduce variability in share time series. This variability is the result of excessive disaggregation – 67 industries and 139 occupations. One solution to this problem is to reduce the industrial dimension or the initial number of occupations projected. The proposed solution addresses both of these. We propose to reduce the industries to 17 and initially project the 2-digit NOC occupations in 25 groups. This way, the initial projection would aim at establishing the broad trends in the employment outlook. Afterwards, the 3-digit occupations would be projected and made to equal the corresponding projections for the 2-digit group. Thus, COPS preserves national projection to 139 NOC 3-digit occupations but in a more aggregate industrial space.

Let us assume a combination of 17 industries and 25 2-digit NOC occupational groups. This would reduce the number of projections to 425 from 9,313, giving us 22 times fewer projections and trends to analyse. Comparison of the series averages and standard deviations for these two samples tells us that the share series of the smaller sample are less volatile than those of the sample containing 9,313 series (see Table 3.1). In fact, the share series of both samples have virtually the same standard deviation, but the average for the 425 series is slightly more than 4 times that of the 9,313 series. The COPS projections would therefore appear to be based on more reliable information by combining industries and making the initial projections with the 2-digit NOC groups.

Table 3.1 Average statistics for share time series

Number of series	Average	Standard deviation
9,313	0.0096	0.0032
425	0.0406	0.0038

3.3 Industrial aggregation: choosing the level and implementation strategy

The reason for greater industrial aggregation is to increase the reliability of the information provided by COPS by eliminating the risk of introducing statistical bias and by concentrating available analytic resources on a smaller number of occupations by industry. Two questions arise here: the appropriate level of aggregation and the best strategy for implementing the change.

With respect to the desired level of aggregation, COPS proposes to use a grouping of 16 industries. Statistics Canada already uses this aggregation level for seasonally adjusted provincial time series.[9] Moreover, when the new classification system, the North American Industry Classification System (NAICS),[10] was developed, Statistics Canada reiterated that a grouping of 16 industries is the minimum aggregation to provide sufficiently reliable seasonally adjusted data at the provincial level.

[9] The monthly provincial breakdown disaggregates the national data 120 times (12 months * 10 provinces), which is close to COPS' disaggregation of 139 occupations by industry.

[10] The purpose of the new NAICS industrial classification is to allow industrial comparisons among North American countries and to make the definitions of industries more consistent and up to date. The new classification groups the branches of activity according to common inputs and production processes, rather than according to outputs, which were the basis of the 1980 SIC. The new system also accommodates new branches of activity that have emerged since the 1970s.

> At the highest level aggregation, NAICS has 20 sectors, compared with 18 divisions in the SIC 1980 classification. Even this level of aggregation is too detailed to provide sufficiently reliable seasonally adjusted estimates for all provinces. Just as was done with SIC 1980, the LFS has regrouped some of the sectors to create 16 for monthly, seasonally adjusted publication.
>
> —Statistics Canada, *Labour Force Update*, Spring 1999, Vol. 3, No. 2, No. 71-005-XPB

With respect to the strategy for implementing the change, COPS first objective is to adopt the NAICS grouping. However, COPS cannot do so immediately because the macroeconomic scenario and the 1996 census are defined according to the 1980 SIC. A macroeconomic scenario based on NAICS should be available in the next few years. The information in the 2001 census will not be available until 2003.

Given the various constraints on immediately adopting NAICS, we propose that the aggregation suggested in Table A1 be adopted on an interim basis while we wait for fully NAICS-based information to be released in 2003.[11] The proposed 1980 SIC grouping differs slightly from the 16 industries used by Statistics Canada: the manufacturing industry is divided in two, durable and non-durable goods, and the fishing and fish processing industries are grouped together because these two industries are closely connected. This grouping is largely similar to the NAICS grouping (see Table A2 of Appendix A).

It will always be possible to disaggregate some or all of these 16 industries again in accordance with COPS' and its partners' needs by using the information from the last census.

[11] This transition could be made earlier – as soon as the macroeconomic scenario based on the NAICS is available and if the 1996 census is converted from the 1980 SIC to NAICS.

4. Second Issue: Taking Economic Information into Account

4.1 Current method versus alternative method

To replace the smoothing equations, alternative equations must as far as possible consider two labour market realities which have a strong influence on the employment series by occupation and by industry: some occupations are substitutes while others are complementary and the economic cycle influences the evolution of certain shares. In other words, the model has to allow shares to vary and must include a component that considers the influence of the economic cycle on certain occupations.[12]

For the purposes of this exercise, two groups of models will be analysed (see Box 4.1). The first group does not incorporate a deterministic trend (equations 1a and 2a) and the second group does (equations 1b and 2b). The equations are written in first difference. We are assuming that first difference makes the occupational share series stationary.[13] This assumption implies that the projection over the 1999-2025 period should not explode.

Box 4.1 Projection equations

Equations without trend	**Equations with trend**
1a. $\Delta X_t = a_1 \Delta X_{t-1} + e_t$	**1b.** $\Delta X_t = a_0 + a_1 \Delta X_{t-1} + e_t$
2a. $\Delta X_t = a_1 \Delta X_{t-1} + a_2 \Delta OG_{t-1} + e_t$	**2b.** $\Delta X_t = a_0 + a_1 \Delta X_{t-1} + a_2 \Delta OG_{t-1} + e_t$

where OG = output gap, deviations from GDP as related to its trend estimated by an HP filter.

[12] The substitution and complementarity factors are present are present in the historical evolution of share time series; however, modelling these effects requires additional research before COPS can integrate this element into its labour demand model.

[13] A series is stationary when its average is constant in time and its variance is independent of the series level. There are specific tests for checking stationarity hypotheses. The large number of series to be projected makes it impossible to conduct all these tests. The small sample (15 years, 1984-98) would also limit the power of these tests.

Equation 1a corresponds to a simple model in which the projected value of an occupational share depends on its past value. The value of coefficient α_1 indicates the persistence level associated with the movement of the estimated share series.[14] When the value of the coefficient approaches 1, there is a lot of temporal persistence in the series. When this value approaches 0, no or few temporal persistence is associated with series movement. We should expect high persistence in movements in an occupational share series that has increased (declined) from period to period. On the contrary: a share series affected by periods of growth and decline would have much less persistence.

Equation 1b incorporates a constant α_0 that corresponds to the average level of the series in first difference. Over the projection horizon, this constant is added to each additional projection period, thus acting similarly to a trend. As a result, the equations including a constant generate occupational share projections that will increase or decrease throughout the projection period.

Equations 2a and 2b are variants of Equations 1a and 1b. The difference stems from the addition of a cyclical component that captures the probable influence of the economic cycle on occupational share evolution. When the coefficient associated with the cyclical variable is equal to zero, Equations 2 produce the same results as Equations 1. When this coefficient differs significantly from zero, the economic cycle is influencing occupational share evolution. This influence is taken into consideration over the entire projection period.[15]

The equations with trends are associated with models where several shares will vary throughout the projection period.[16] The equations without trends will produce projections halfway between a

[14] The persistence level determines the velocity at which the growth of the projected share converges on 0. Convergence velocity corresponds to the number of periods that elapse before the projected series stops growing on the projection horizon. The greater the persistence, the slower convergence on zero will be, and several projection periods will be needed before the occupational share stops growing. For example, if the coefficient α1 was equal to 0.7, the share would take something over ten years to become practically fixed. With a coefficient of 0.7, the effect equals 0.7 in the first projection period, in the second period the effect is 0.49 or 0.7 * 0.7 or 0.72, and so forth. The added effect on share growth in the tenth projection period will be 0.7^{10} or 0.03. With a coefficient α_1 of 0.3, the share would take only three or four years before ceasing to evolve over the projection period. This natural process of convergence enables us to eliminate the manual intervention of imposing a growth ceiling or floor (see Stage 2 c in Box 2.1).

[15] We are using the historical and projected GDP in the macroeconomic scenario for calculating the cyclical component between 1984 and 2008.

[16] When the trend is insignificant, the share time series will be more or less fixed over the projection period.

variable share model and a fixed share model. The length of a share's period of evolution is directly related to that share's persistence level. When persistence is high, the share will evolve for several periods and become fixed thereafter. This behaviour is consistent with the idea that in the medium term we can claim partial knowledge of the direction in which a series may evolve. However, in the long term it would be pretentious to think that we possess this information. It is therefore prudent to maintain the long-term fixed share.

4.2 Evaluation of predictive performance at the national level

The current methodology and the four proposed equations were estimated with LFS data for the period 1984-96 and projected for the years 1997 and 1998.[17] For the purposes of evaluating the current and proposed equations, our estimates use the occupational total (no industries). The choice of having no industries stems from a desire to conduct the evaluation with series containing the least possible statistical noise. This consideration is important to the robustness of our evaluation results.

These equations were initially estimated and projected with each of the 25 2-digit occupational groups in the NOC. Afterwards, we derived the projection for the 3-digit NOC groups. The methodology used with the proposed equations to obtain the 3-digit NOC projections based on the 2-digit NOC projections is explained in Appendix B.

The evaluation is based on five criteria:[18] the mean error projection (MEP), the Theil inequality coefficient, the BIAS that indicates the systematic projection error, the VARIANCE that indicates the model's ability to reproduce the variability of the series observed, and the proportion of the COVARIANCE that measures the unsystematic error, the part which cannot be attributed to the BIAS and the VARIANCE. The first four criteria are to be minimized and the last one maximized.

[17] This choice is justified inasmuch as we have very few observations for making our estimate. A longer period would have afforded us a longer projection period for evaluation purposes.
[18] The statistical criteria used to evaluate the equations are set out in Appendix C.

The evaluation results presented in Table 4.1 indicate that the four proposed equations generally obtain better evaluation results than the ones obtained with the current methodology. The first three criteria obtain better results with the proposed equations. However, the current methodology obtains the best result with the variance and it is impossible to discriminate between the results of the last criterion.

Table 4.1 Projection evaluation criteria, average for the 139 3-digit NOC occupations, total for industries

	EMP	THEIL	BIAS	VARIANCE	COVA
Actual equation	0.00028 (5)	0.04224 (5)	0.63191 (5)	0.23579 (1)	0.13230 (4)
1a.	0.00018 (2)	0.02904 (2)	0.50000 (2)	0.37941 (5)	0.12059 (5)
1b.	0.00019 (4)	0.03077 (4)	0.50000 (2)	0.33894 (2)	0.16106 (1)
2a.	0.00017 (1)	0.02851 (1)	0.49999 (1)	0.36166 (4)	0.13834 (3)
2b.	0.00018 (3)	0.03007 (3)	0.50000 (3)	0.34697 (3)	0.15303 (2)

Note: Estimation period 1984-96, evaluation period 1997-98

Based on these evaluation criteria, it appears that the proposed equations provide better projections than those obtained with the method currently used by COPS. Of the proposed equations, the equation with no linear trend and a cyclical component (2a) provides the best evaluation results based on the first three criteria. Further, this equation takes into account the issue associated with the influence of the economic cycle on certain occupations. In the next section, we shall evaluate the implications of Equation 2a for the performance of provincial projections.

4.3 Evaluation of predictive performance at the provincial level

For the purposes of this evaluation, the predictive performance of Equation 2a at the provincial level will be assessed using only 3-digit NOC occupations, because the 4-digit NOC occupations are the distribution of the corresponding 3-digit NOC family. It is important to note that the proposed change in methodology entails no change in the level of detail; COPS would continue to produce provincial projections for 511 4-digit NOC occupations.

4.3.1 Deriving provincial projections from national projections: review of the current methodology

To obtain provincial industrial and occupational employment projections, COPS uses three components: the national occupational share projection by industries (3-digit NOC), the provincial occupational share by industries according to the 4-digit NOC and the employment of the provincial macro industrial scenarios. The provincial projections are derived in three stages:

1. **Creation of 4-digit NOC occupational share series by industries for each province**: adjusting the level of the national projected 3-digit NOC share series to that of the provincial occupational share corresponding to the 4-digit NOC from the 1996 provincial census.
2. **Normalization**: the 4-digit provincial shares obtained in Stage 1 are normalized in order to guarantee that the sum of shares equals 1 within each industry.
3. **Provincial employment by occupation and industry**: multiplication of the 4-digit provincial share series by the corresponding province's macro industrial employment scenario.

It is important to note that the trend in provincial 3-digit shares is exactly the same as the national 3-digit trend. To a large extent, then, the provincial and national projections behave quite similarly. The differences in industrial occupational employment that do exist between national and provincial projections stem from the use of macro industrial scenarios and 4-digit provincial share levels that vary from province to province.

4.3.2 Evaluation of projections

Three provincial projection methods are evaluated in this section: the current smoothing method and Equation 2 a) from Section 4, estimated separately with national series and with provincial output gap series. Comparison of the predictive performance of the equations using different output gap series will seek to determine whether estimating and projecting *national* shares with *provincial* output gaps will introduce an additional bias into the projection. If the performance of the two equations is comparable, the provincial output gap series will be used, which will introduce a third variability factor specific to each province, the two others being the weight of the census and the use of the provincial macro industrial scenario. These three methods are estimated for the period 1984-96 and projected for the years 1997 and 1998.

Table 4.2 Projection evaluation criteria, average for the 139 3-digit NOC occupations, total for industries

	MEP	THEIL	BIAS	VARIANCE	COVA
Newfoundland					
Current (smoothing)	0.00061	0.08983	0.74735	0.09081	0.16184
National (national GDP)	0.00037	0.04488	0.67563	0.11531	0.20905
Provincial (provincial GDP)	0.00036	0.04586	0.67740	0.11507	0.20753
Prince Edward Island					
Current (smoothing)	0.00063	0.09560	0.72799	0.08111	0.19090
National (national GDP)	0.00036	0.05209	0.63158	0.11713	0.25129
Provincial (provincial GDP)	0.00035	0.05357	0.61678	0.12990	0.25332
Nova Scotia					
Current (smoothing)	0.00046	0.06259	0.74421	0.08367	0.17212
National (national GDP)	0.00035	0.04236	0.65881	0.11124	0.22995
Provincial (provincial GDP)	0.00036	0.04462	0.66005	0.10630	0.23365
New Brunswick					
Current (smoothing)	0.00054	0.07185	0.73030	0.08797	0.18173
National (national GDP)	0.00035	0.04191	0.65364	0.11281	0.23355
Provincial (provincial GDP)	0.00037	0.04408	0.64305	0.12119	0.23576
Quebec					
Current (smoothing)	0.00038	0.05029	0.68119	0.11212	0.20669
National (national GDP)	0.00036	0.04295	0.64975	0.11135	0.23889
Provincial (provincial GDP)	0.00036	0.04361	0.65106	0.10783	0.24112
Ontario					
Current (smoothing)	0.00038	0.05054	0.66436	0.11667	0.21897
National (national GDP)	0.00036	0.04444	0.63067	0.11970	0.24963
Provincial (provincial GDP)	0.00038	0.04594	0.64323	0.11335	0.24342
Manitoba					
Current (smoothing)	0.00041	0.05909	0.71643	0.09949	0.18408
National (national GDP)	0.00034	0.04250	0.64326	0.11980	0.23694
Provincial (provincial GDP)	0.00038	0.04522	0.64976	0.11675	0.23349
Saskatchewan					
Current (smoothing)	0.00043	0.06678	0.72911	0.08524	0.18565
National (national GDP)	0.00032	0.04311	0.66932	0.10683	0.22385
Provincial (provincial GDP)	0.00033	0.04179	0.65820	0.10966	0.23214
Alberta					
Current (smoothing)	0.00042	0.06014	0.71371	0.08870	0.19760
National (national GDP)	0.00034	0.04570	0.66501	0.09801	0.23698
Provincial (provincial GDP)	0.00035	0.04593	0.65412	0.09908	0.24680
British Columbia					
Current (smoothing)	0.00046	0.05891	0.71618	0.10147	0.18235
National (national GDP)	0.00035	0.04259	0.65416	0.10862	0.23722
Provincial (provincial GDP)	0.00035	0.04393	0.67954	0.09914	0.22131

Provincial projections are evaluated with the same criteria as the ones used in Section 4.1. To obtain the provincial share projection, the projection of the three methods is initially multiplied in the provincial macro industrial employment scenarios, which generates provincial employment by occupation and industry.[19] After this, total employment by occupation is calculated as the sum for all COPS industries. Finally, the projected shares are obtained by dividing total employment in each occupation by total provincial employment. These projections are compared with the provincial history, calculated by multiplying national shares in provincial macro industrial employment scenarios, then adding employment by occupation for all industries and dividing the total for each occupation by total provincial employment to obtain the historical provincial share.

The results confirm that the new method obtains better evaluation results than those obtained with the current smoothing method (see Table 4.2). However, the evaluation results do not enable us to differ between the new method estimated with the national series and the one estimated with the provincial output gap series. The predictive performance of the two equations is very comparable. Provincial output gap series will therefore be used to estimate and project the 3-digit NOC shares series at the provincial level, which will introduce a third variability factor specific to each province.

[19] In evaluating national projection methods, we have not used industries since, at this time, our concern was to determine which method afforded the better projection performances. However, we did use industries in the provincial evaluation to clearly bring out the implications of using the national industrial output gap series in Equation 2 a compared to the provincial industrial output gap series.

5. Comparisons of Projections: Current Versus Proposed Method

In the preceding section, based on statistical criteria, we showed that the proposed methodology allows for better forecasting. In this section, we compare the projections obtained using the current and proposed methodologies for 1999-2008. The two methodologies use a different framework to estimate and project the economic signal from share series. The current methodology uses smoothing statistical techniques that arbitrarily reallocate employment by occupation over the historical period. The proposed methodology models the economic information contained in the historical behavior of share series to project future trends more effectively.

Despite the methodological differences that exist between the two approaches, the projections obtained follow the same trend but at different growth rates. The fact that the proposed methodology maintains the historical behavior of share series allows us to predict that the resulting projections should better reflect the future trend of share series.

The growth of expansion demand and the average annual growth rate are compared between the two methodologies over three five-year periods: the most recent historical period (1994-1998) and two projection periods (1999-2003 and 2004-2008). The results for the five skill levels are shown in Figure and Table 5.1. The comparison of the 25 occupational groups in the 2-digit NOC is shown in Figure and Table A4.1 in Appendix D.

Between 1994 and 1998, the growth of expansion demand for managers is 51,000 jobs higher with the current methodology compared with the proposed methodology. This difference is entirely attributable to the current methodology, which arbitrarily redistributes employment over the historical period. The difference in the projection of the growth of demand for this group (45,000 jobs in 1999-2003 and 56,000 in 2004-2008) is chiefly explained by the difference in behavior over the historical period. The proposed methodology interprets the plateauing of the growth of employment in 1997-1998 as a slowdown in the future growth of occupational shares of groups comprising the management group. The current methodology, on the other hand, completely misses this slowdown and produces a projection that behaves as if this event had never occurred. Specifically, the current methodology estimated that average annual growth for managers would total 2.3% in 1994-1998 and 1999-2003, compared with 1.7% and 1.9% for the corresponding periods with the proposed methodology.

Table 5.1 Projection comparison: current method vs. proposed method

	Expansion demand (000s)			Average annual growth rate		
	1994-98	1999-03	2004-08	1994-98	1999-03	2004-08
Managers						
Current	168.5	188.7	168.8	2.3	2.3	1.9
Proposed	117.8	143.9	112.6	1.7	1.9	1.3
Difference	50.7	44.8	56.2	0.6	0.4	0.6
Skill-Level A						
Current	334.2	271.0	221.1	3.2	2.2	1.7
Proposed	345.5	166.9	172.2	3.3	1.4	1.4
Difference	-10.3	114.2	48.9	-0.1	0.8	0.3
Skill-Level B						
Current	216.0	355.7	325.7	1.0	1.6	1.4
Proposed	282.3	371.8	320.0	1.3	1.6	1.3
Difference	-66.3	-16.1	5.7	-0.3	0.0	0.1
Skill-Level C						
Current	406.2	265.6	267.5	1.9	1.2	1.1
Proposed	408.8	340.6	341.9	2.0	1.5	1.4
Difference	-2.6	-75.0	-74.4	-0.1	-0.3	-0.3
Skill-Level D						
Current	185.0	97.3	111.2	2.3	1.1	1.2
Proposed	147.5	146.9	139.2	1.8	1.7	1.5
Difference	37.5	-49.6	-28.0	0.5	-0.6	-0.3

For skill-level A occupations, the difference in expansion demand between the two methodologies totals 114,000 and 49,000 respectively in 1999-2003 and 2004-2008. These differences stem mainly from the Professional Occupations in Business and Finance group (Group 11) and the Professional Occupations in Natural and Applied Sciences group (Group 21). In the case of professional occupations in business and finance, a large share of the difference in the two projections stems from the difference in the historical behavior of the share series. The proposed methodology considers observable cyclical movements in the early 1990s. This means a cyclical adjustment at the beginning of the projection period, which then affects the trend and growth rate of the share series. Among professional occupations in natural and applied sciences, the share series experienced strong growth between 1996 and 1998, caused largely by the Y2K significant upward adjustment of the trend estimated and extrapolated by the current methodology. However, both methodologies produce an upward projection in employment for professional occupations in natural and applied sciences, but at different growth rates. The growth produced by the proposed methodology approaches the behavior observed in 1984-1995.

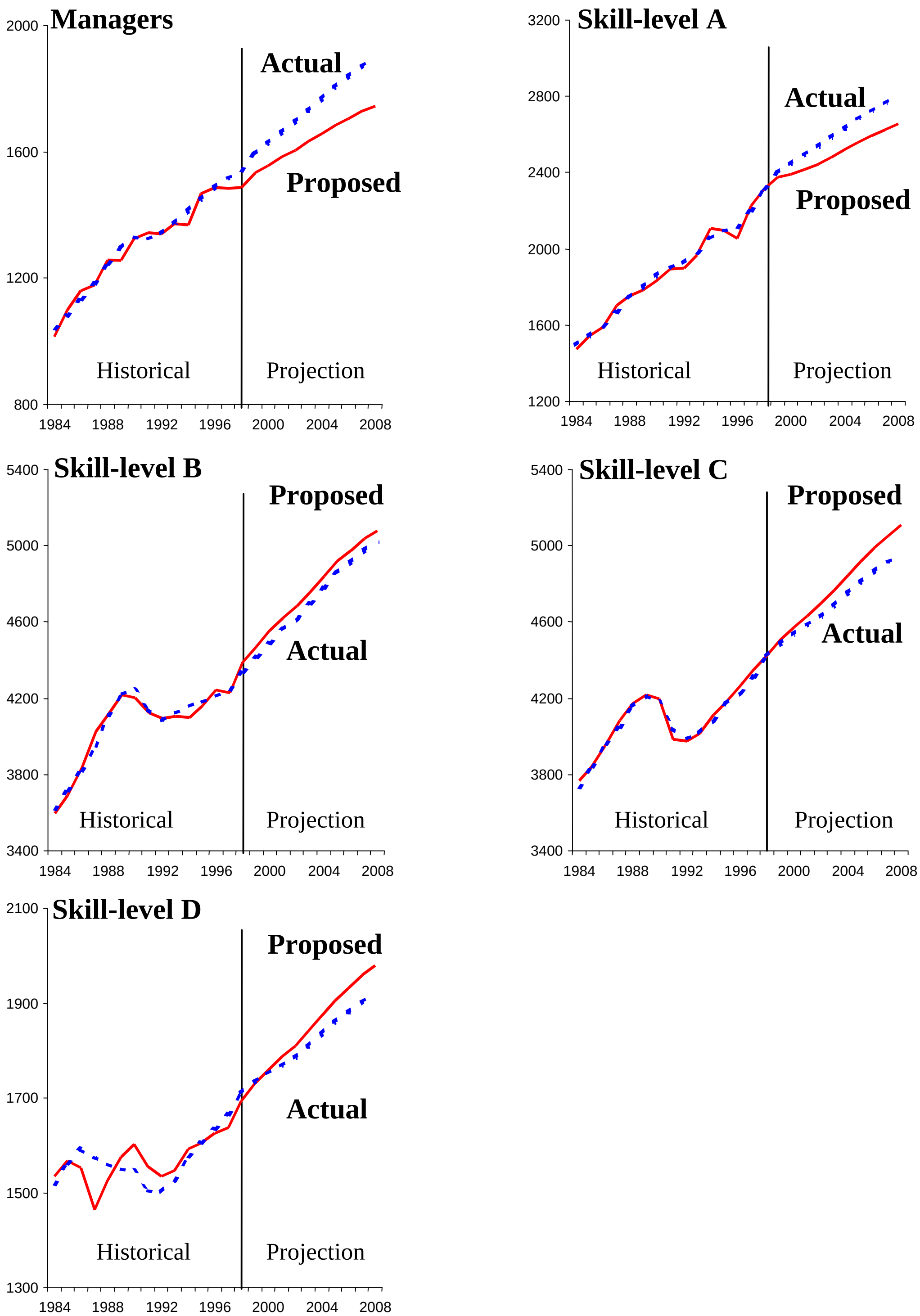
The projection of skill-level B occupations obtained with the two methodologies yields very similar results. However, there are differences within this skill level over the projection period. The proposed methodology produces stronger growth (1.0% and 0.4%) than the current methodology for skilled administrative and business occupations (Group 12) and slower growth (-0.3% and -0.4%) for skilled occupations in primary industry (Group 82).

The projected expansion demand for skill-level C occupations is 75,000 jobs higher in 1999-2003 and 2004-2008 with the proposed methodology than with the current methodology. This difference stems almost entirely from the projection for clerical occupations (Group 14). Part of this difference is explained by different growth over the historical period (-0.3% with the proposed methodology compared with -0.8% with the current methodology in 1994-98). The remainder of this difference stems from how the economic signal is treated by the proposed methodology, which sharply slows down the negative growth of this share of the group over the projection period. The current methodology, on the other hand, continues the negative growth of this share at a faster rate.

According to the current methodology, the average growth of skill-level D occupations will decrease from 2.3% in 1994-1998 to 1.1% and 1.2% over the two projection periods, compared with growth rates of 1.8%, 1.7% and 1.5% for the same periods with the proposed methodology. Once again, historical differences result in different projections. The main difference stems from the Elemental Sales and Service Occupations group (Group 66). The proposed methodology projects a slight increase in the employment share for this group, compared with a decrease in the employment share with the current methodology.

In summary, the use of different methodological framework explains the differences in behavior observed over the historical period and the projection period from the two approaches. The fact that the proposed methodology maintains the historical behavior of share series and models economic information allows us to predict that the resulting projections will better reflect the future trend of job series by occupation and industry.

Figure 5.1 Jobs (000) by skill level



6. Conclusions and Future Developments

This analysis attempted to answer the following two questions:

- Does the current methodology enable us to fully take into consideration the economic information contained in the occupational series?
- Is the current level of disaggregation, 67 industries by 139 occupations, appropriate in terms of minimizing the risk of introducing a bias into projections given the small size of the Labour Force Survey (53 000 households)?

With respect to the first question, the analysis has shown that the methodology currently used by COPS, smoothing, did not enable us to fully take into consideration the economic information contained in occupational employment series by industries. A new methodology was therefore proposed and evaluated in Section 4. Analysis of the new models' predictive performance shows that an equation that includes a cyclical component produces better projections than the smoothing equation used by COPS at both the national and provincial levels. It was also demonstrated that the proposed model better reflects the scope of economic trends that govern the behavior of some occupational and industrial share series, without arbitrarily modifying the series' historic behavior.

In answer to the second question, the analysis suggests that the current disaggregation level is not appropriate to minimize the risk of introducing a bias into projections. The current COPS method produces 9,313 share projections. The information for these share time series comes from the LFS, a Statistics Canada survey where the sample consists in barely 53,000 households or slightly under 6 households per occupational employment series by industries. These series are therefore affected by a high statistical noise level that prompted the introduction of smoothing by COPS. To correct this situation, the analysis suggests reducing the industrial component to 17 industries and initially projecting the occupational shares of the 25 major 2-digit NOC groups, then the 3-digit NOC groups. In this way, the 425 initial projections (17 industries and 25 occupational groups) would aim to establish the broad trends in the employment outlook and cut the number of trends to be analysed 22 times. The comparative statistics of these two samples (9,313 series vs. 425) support our aggregation approach.

This analysis leads to the following recommendations for the next COPS occupational projections:

1. Use an equation with no linear trend but with a component that takes into consideration the probable influence of the economic cycle and temporal persistence (Equation 2a) to replace the current COPS share projection method.

2. Reduce the number of occupations in the initial projection phase in order to establish the broad trends in the employment outlook and reduce the industrial component to lower the statistical noise level present in share time series.

The recommended changes would increase the sensitivity of the share projections while providing the same level of detail in the occupational demand projections as does the current methodology.

To model the existing interactions between certain occupations, COPS will embark on research for a better understanding of labour market occupational and industrial segmentation. This research will seek to improve our understanding of the similarities and differences in behavior between certain occupations, and of inter-occupational and inter-industrial mobility inside and between occupational groups.

Appendix A

Proposed Industrial Aggregation

Table A1 Proposed aggregation	Number of COPS industries (67)
1: Agriculture	1
2: Fishing and trapping	2, 10
3: Forestery	3
4: Mining	4, 5, 6, 7, 8, 9
5: Non-durable goods	11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23
6: Durable goods	24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36
7: Construction	37
8: Transportation	38, 39, 40, 41, 42, 43, 44
9: Communications	45, 46, 47
10: Public utilities	48, 49, 50
11: Business	51, 52
12: Finance, insurance, real estate	53
13: Professional services	54, 55, 56
14: Public administration	57, 58, 59
15: Teaching	60
16: Health services	61, 62, 63
17: Sales and personal services	64, 65, 66, 67

Table A2 Sixteen industries based on North American Industrial Classification Systems
1: Agriculture
2: Forestry, fishing, mining, oil and gas
3: Utilities
4: Construction
5: Manufacturing (can be divided in durable and non durable goods)
6: Trade
7: Transportation and warehousing
8: Finance, insurance, real estate and leasing
9: Professional, scientific and technical services
10: Management, administrative and other support
11: Educational services
12: Health care and social assistance
13: Information, culture and recreation
14: Accommodation and food
15: Other services
16: Public administration

Appendix B

Three-digit NOC Share Projections Based on 2-digit NOC Groups

To derive the 3-digit occupational share projections from 2-digit occupational share projections, the 3-digit occupations are considered as being the distribution of the corresponding 2-digit occupations. After this, these shares are projected to the projection horizon with the selected equation. Lastly, the projected 3-digit shares are applied to the 2-digit occupational share projection. This compels the sum of the 3-digit share projections to equal the 2-digit share projection of which they are the components.

As an example, S11 is made up of S111, S113 and S117 where S11 = S111 + S113 + S117. To begin with, the 3-digit occupational shares are made into shares that add up to 1 as follows:

$$PS111_t = {}^{1}\!/_{S11_t} * S111_t$$

$$PS113_t = {}^{1}\!/_{S11_t} * S113_t$$

$$PS117_t = {}^{1}\!/_{S11_t} * S117_t$$

where $PS111_t + PS113_t + PS117_t = 1$

These shares are afterwards projected with the selected equation and noted down as $PS111F$, $PS113F$ and $PS117F$. The next step is to adjust the projected shares to make their sum equal to 1. Finally, the 3-digit share projections are multiplied by the 2-digit occupational share projection as follows:

$$S111F_t = PS111F_t * S11F_t$$

$$S113F_t = PS113F_t * S11F_t$$

$$S117F_t = PS117F_t * S11F_t$$

where $S11F$ corresponds to the 2-digit projection.

Appendix C

Projection Evaluation Criteria

The first criterion measures deviations in the projected series relative to the observed series. This measurement[20] is called the root of the mean forecasting error (*RMFE*) and is defined as follows:

$$EMP = \sum_{s=CTP} a_s \sqrt{\frac{1}{T}\sum_{t=1}^{T}\left(Y_{st}^{p} - Y_{st}^{a}\right)^2} \tag{A.3.1}$$

where Y_t^p = the projected series, Y_t^a = the current series and T = the number of projection periods. The weight a_s corresponds to the average share for each occupation and is used to obtain an aggregated measurement of the performance of each equation for all occupations.

The second measurement is the *Theil* inequality coefficient and is defined as follows:

$$Theil = \sum_{s=CTP} a_s \frac{\sqrt{\frac{1}{T}\sum_{t=1}^{T}\left(Y_{st}^{p} - Y_{st}^{a}\right)^2}}{\sqrt{\frac{1}{T}\sum_{t=1}^{T}\left(Y_{st}^{p}\right)^2} + \sqrt{\frac{1}{T}\sum_{t=1}^{T}\left(Y_{st}^{a}\right)^2}} \tag{A.3.2}$$

Note that the numerator is the *MEP* and the scale of the denominator is such that the value of the statistic will always be between 0 and 1. When this statistic equals 0, $Y_t^p = Y_t^a$. However, a value equal to 1 would indicate the worst possible projection. The *RMFE* can be broken down into three parts: *bias, variance* and *covariance*. These measurements are written as follows:

$$biais = \sum_{s=CTP} a_s \frac{\left(\overline{Y}_s^{\,p} - \overline{Y}_s^{\,a}\right)^2}{(1/T)\sum_{t=1}^{T}\left(Y_{st}^{p} - Y_{st}^{a}\right)^2} \tag{A.3.3}$$

[20] For more details about this and the next four measurements, see Chapter 12 of Pindyck and Rubinfeld (1991).

$$variance = \sum_{s=CTP} a_s \frac{\left(s_p - s_a\right)^2}{\left(1/T\right)\sum_{t=1}^{T}\left(Y_t^p - Y_t^a\right)^2} \tag{A.3.4}$$

$$covariance = \sum_{s=CTP} a_s \frac{2\left(1 - r\right)s_p s_a}{\left(1/T\right)\sum_{t=1}^{T}\left(Y_{st}^p - Y_{st}^a\right)^2} \tag{A.3.5}$$

where $\bar{Y}$ indicates the average value of the projected and observed series, σ_p and σ_a are the standard deviations of each series and ? is their correlation coefficient.

The proportion of the *bias* is an indication of the systematic projection error. It measures the degree to which the average values of the projected and observed series deviate from one another. The proportion of the *variance* indicates the model's ability to reproduce the variability of the observed series. Finally, the proportion of the *covariance* measures the unsystematic error that cannot be attributed to the *bias* and the *variance*. To determine which model generates the best projections, we must choose the one showing the smallest values for the first four criteria while being the closest to 1 for the fifth criterion.

Appendix D

Comparison of Projections by 2-digit NOC Groups

Table D.1: Comparison of projections by major 2-digit NOC groups: current vs. proposed methodology

		Expansion demand (000)			Average annual growth rate		
		1984-98	1999-03	2004-08	1984-98	1999-03	2004-08
GR 00	Current	168.5	188.7	168.8	2.3	2.3	1.9
	Proposed	117.8	143.9	112.6	1.7	1.9	1.3
	Difference	50.7	44.8	56.2	0.6	0.4	0.6
GR 11	Current	98.1	70.1	45.5	5.6	3.2	1.8
	Proposed	119.6	34.4	30.2	7.2	1.6	1.3
	Difference	-21.5	35.7	15.3	-1.6	1.6	0.5
GR 12	Current	-87.7	31.4	39.6	-1.9	0.7	0.9
	Proposed	-38.7	78.3	64.2	-0.8	1.7	1.3
	Difference	-49.0	-46.9	-24.6	-1.1	-1.0	-0.4
GR 14	Current	-46.0	-3.3	11.3	-0.8	-0.1	0.2
	Proposed	-19.7	83.6	79.2	-0.3	1.4	1.2
	Difference	-26.3	-86.9	-67.9	-0.5	-1.5	-1.0
GR 21	Current	138.8	79.0	61.5	6.6	2.9	2.0
	Proposed	138.4	44.2	40.8	6.5	1.7	1.4
	Difference	0.4	34.8	20.7	0.1	1.2	0.6
GR 22	Current	43.8	26.0	19.0	3.1	1.6	1.1
	Proposed	44.2	21.5	23.1	3.1	1.3	1.3
	Difference	-0.4	4.5	-4.1	0.0	0.3	-0.2
GR 31	Current	12.1	37.2	34.8	0.6	1.8	1.6
	Proposed	17.8	17.4	31.6	0.9	0.9	1.5
	Difference	-5.7	19.8	3.2	-0.3	0.9	0.1
GR 32	Current	2.5	18.1	14.3	0.3	1.9	1.4
	Proposed	-6.8	27.2	16.4	-0.8	2.9	1.6
	Difference	9.3	-9.1	-2.1	1.1	-1.0	-0.2
GR 34	Current	25.5	19.4	19.1	3.2	2.1	1.9
	Proposed	18.6	22.0	15.3	2.3	2.5	1.5
	Difference	6.9	-2.6	3.8	0.9	-0.4	0.4
GR 41	Current	59.6	66.6	56.4	1.5	1.6	1.2
	Proposed	57.8	56.9	52.5	1.5	1.4	1.2
	Difference	1.8	9.7	3.9	0.0	0.2	0.0
GR 42	Current	37.3	20.5	16.5	4.4	2.0	1.5
	Proposed	34.1	16.2	15.9	4.1	1.7	1.5
	Difference	3.2	4.3	0.6	0.3	0.3	0.0
GR 51	Current	25.6	18.1	23.0	3.0	1.9	2.2
	Proposed	12.0	14.1	17.1	1.4	1.5	1.7
	Difference	13.6	4.0	5.9	1.6	0.4	0.5
GR 52	Current	44.0	14.0	18.2	4.6	1.2	1.5
	Proposed	39.8	18.1	20.3	4.2	1.6	1.7
	Difference	4.2	-4.1	-2.1	0.4	-0.4	-0.2

Table D.1 (Continued)

		Expansion demand (000)			Average annual growth rate		
		1984-98	1999-03	2004-08	1984-98	1999-03	2004-08
GR 62	Current	36.6	103.1	99.5	1.0	2.5	2.1
	Proposed	60.6	60.2	68.1	1.6	1.5	1.5
	Difference	-24.0	42.9	31.4	-0.6	1.0	0.6
GR 64	Current	188.1	144.7	129.7	2.7	1.9	1.5
	Proposed	165.4	126.6	131.8	2.4	1.6	1.6
	Difference	22.7	18.1	-2.1	0.3	0.3	-0.1
GR 66	Current	119.9	85.1	99.9	2.0	1.3	1.5
	Proposed	83.7	120.1	113.4	1.4	1.9	1.6
	Difference	36.2	-35.0	-13.5	0.6	-0.6	-0.1
GR 72	Current	138.8	109.0	90.2	2.3	1.7	1.3
	Proposed	127.7	126.9	95.0	2.1	1.9	1.3
	Difference	11.1	-17.9	-4.8	0.2	-0.2	0.0
GR 74	Current	94.4	62.0	61.1	2.9	1.7	1.5
	Proposed	93.0	58.4	50.4	2.8	1.6	1.3
	Difference	1.4	3.6	10.7	0.1	0.1	0.2
GR 76	Current	0.5	1.7	-1.4	0.1	0.3	-0.2
	Proposed	4.7	15.5	8.8	0.8	2.4	1.2
	Difference	-4.2	-13.8	-10.2	-0.7	-2.1	-1.4
GR 82	Current	2.3	22.5	13.5	0.1	1.2	0.7
	Proposed	2.1	16.8	5.1	0.1	0.9	0.3
	Difference	0.2	5.7	8.4	0.0	0.3	0.4
GR 84	Current	-9.2	-7.7	-8.6	-1.2	-1.0	-1.2
	Proposed	-14.5	5.1	1.8	-1.7	0.6	0.2
	Difference	5.3	-12.8	-10.4	0.5	-1.6	-1.4
GR 86	Current	15.0	3.0	0.6	4.1	0.7	0.1
	Proposed	9.5	3.5	3.7	2.4	0.8	0.8
	Difference	5.5	-0.5	-3.1	1.7	-0.1	-0.7
GR 92	Current	-1.6	11.0	14.9	-0.2	1.5	1.9
	Proposed	19.4	6.6	11.9	2.9	0.9	1.5
	Difference	-21.0	4.4	3.0	-3.1	0.6	0.4
GR 94	Current	153.6	50.5	54.9	5.0	1.4	1.4
	Proposed	166.0	44.8	63.5	5.5	1.2	1.6
	Difference	-12.4	5.7	-8.6	-0.5	0.2	-0.2
GR 96	Current	49.7	7.5	12.1	4.5	0.6	0.9
	Proposed	49.6	7.7	13.3	4.5	0.6	1.0
	Difference	0.1	-0.2	-1.2	0.0	0.0	-0.1

Figure D.1 Comparison of employment (000) projections by major 2-digit NOC groups: current vs. proposed methodology

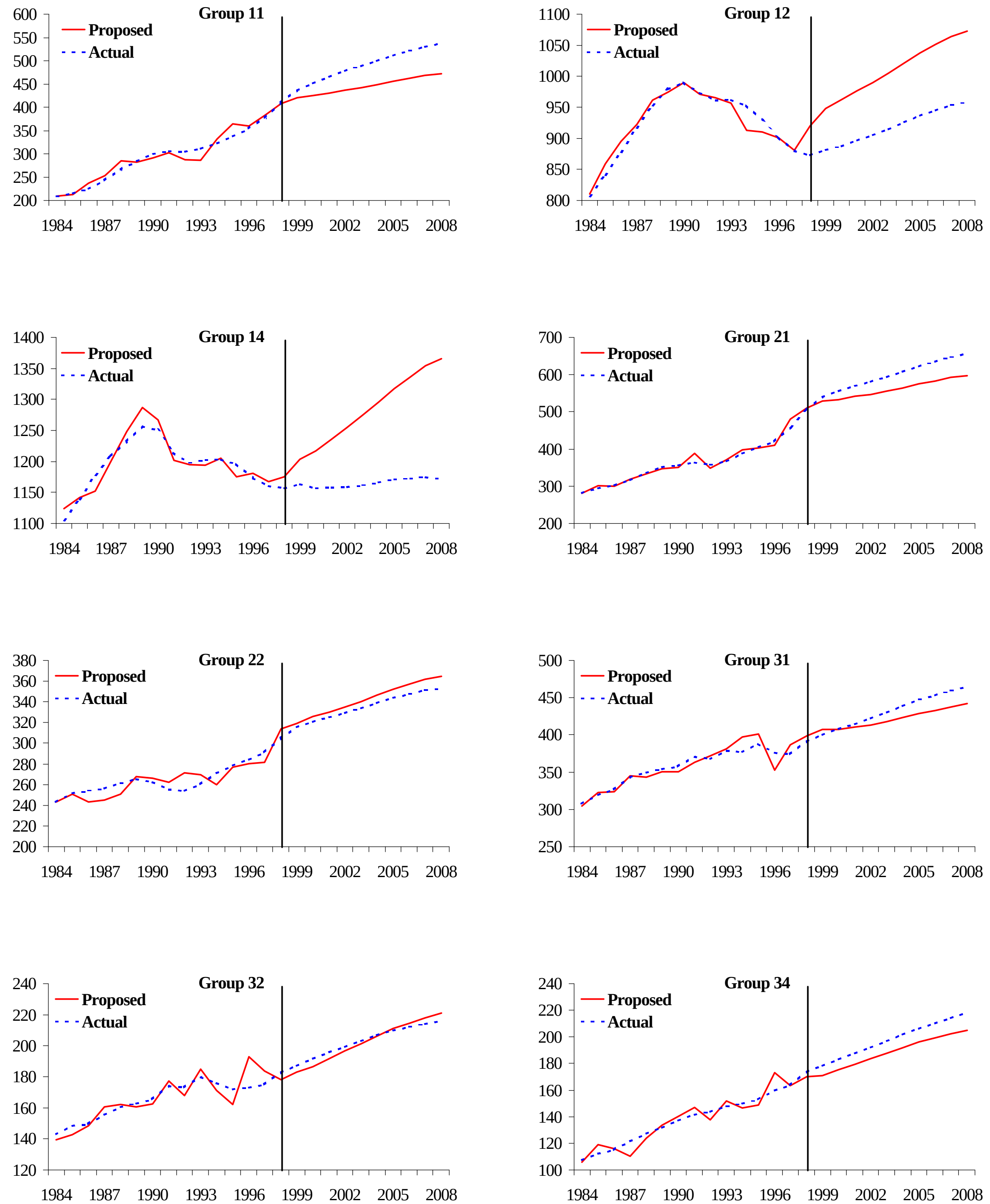


Figure D.1 (Continued)

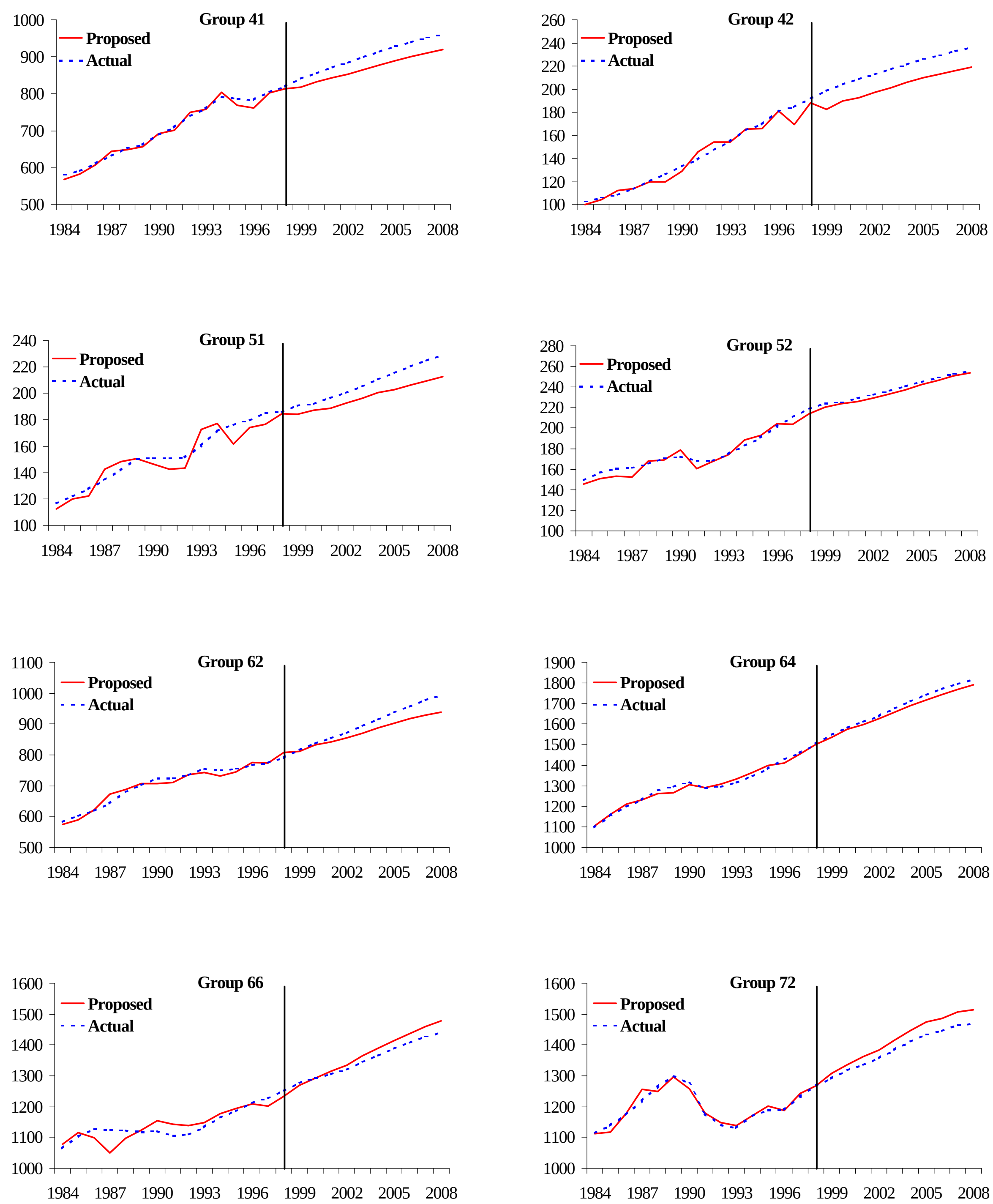


Figure D.1 (Continued)