10 July 2017

ENGLISH ONLY

International Civil Service Commission
Eighty-fifth session
Vienna, 10 July – 21 July 2017
Item 5(c) of the provisional agenda
Conditions of service of the Professional and higher categories:
Post adjustment matters

Considerations regarding cost-of-living surveys and post adjustment matters

Note by Geneva-based organizations[1]

Introduction and Background

1. Comprehensive cost-of-living surveys were conducted by the International Civil Service Commission (ICSC) at the seven headquarters duty stations outside New York (Geneva, London, Madrid, Montreal, Paris, Rome and Vienna) and Washington, D.C. in September/October 2016 for the purpose of gathering price and expenditure data to be used for the determination of the post adjustment index at those locations. In the years prior to this round of surveys, the ICSC had approved a number of changes to the survey methodology based on recommendations of the Advisory Committee on Post Adjustment Questions (ACPAQ). The full set of recommendations brought a significant number of changes to the methodology underlying the 2016 round of surveys, the most notable of them being:

a. The adoption of the European Comparison Programme (ECP) as a source of specifications for the updated basket of goods and services, except for items whose prices were to be collected by the survey coordinator, mostly services and electronic or technological items subject the Real Time Price Comparisons (RTPC) approach, or directly by the secretariat, cars and scooter, also subject to the RTPC approach;

b. The adoption of the ECP as a source of price data for a large share of the basket of goods and services for ECP covered duty stations (London, Madrid, Paris Rome and Vienna). On the other hand, this methodological change did not affect all headquarters duty stations, as for Geneva, Montreal, New York and Washington D.C., price data were collected directly by the ICSC secretariat, in collaboration with the Local Survey Committees (LSC), in accordance with the approved guidelines;

[1] **This document is presented without prejudice to the organizations' position in the event of legal challenge in connection with any of the matters under consideration.**

c. The change in the calculation of average prices, whereas the prior two-step approach of deriving first within-store averages and secondly, determining the item average price as an across-outlets simple average was discontinued and replaced by a simple average of price quotations across all surveyed outlets in one step only;

d. The design of staff expenditures survey questionnaires, combining the two separate questionnaires on housing and domestic services costs and household expenditures, used in prior rounds of surveys, into one questionnaire, the aggregation of questions on expenditures in the food, non-alcoholic beverages and alcoholic beverages categories, the aggregation of questions of formerly distinct basic headings and the introduction of new or more disaggregated questions in other areas of the questionnaire. This allowed for the elimination of repetition and redundancies, and considerably reduced the response burden by reducing the number of questions asked to respondents;

e. The change in the basic heading structure, based on a further reduced number of basic headings, to 80 from the 84 used in the 2010 round, and the procedure of estimating those selected subsets of the common weights related to basic headings of the In-Area (excluding Housing) component grouped under one question only (while maintaining the earlier criteria regarding the required accuracy of the estimates, set by the Committee as coefficients of variation of at most 15 percent, for each grade at a given duty station, and 25 per cent across all grades and duty stations, and post-stratification by collapsing estimates across duty stations in case results were not acceptable for some duty stations, and resorting to the use of external data only as a last resort);

f. The change in the specification of the Out-of-Area (OA) weight, now better reflecting the reality as measured by the survey than the administrative formula used universally for group I duty stations in the 2010 round, which set the applicable OA weight as 20 percent of net remuneration of a staff member at the level of P-4 Step 6 with dependents (if the actual OA weight is indeed less than 20 percent of net remuneration) plus Non Consumption Commitments (NCC), estimated as 5 percent of the net base salary;

g. The new weighting pattern of the 26 countries whose consumer price indices (expressed in US dollars) are used in the calculation of the OA Index, determined on the basis of a special global staff survey conducted in 2012;

h. An updated set of longevity weights for the calculation of the 6-year moving average of rents, and duty-station specific patterns of dwelling type and size weights used in the calculation of the rent index for group I duty stations;

i. The use of an expanded set of New York average rent levels, as surveyed by the International Service for Remunerations and Pensions (ISRP), covering not only the traditional and most typical living surface areas of the target dwelling types and sizes, but also new and less typical ones, to match the dwelling types and sizes of other group I duty stations, thereby rendering unnecessary the use of dwelling size adjustment factors; and

j. The expansion of the staff population eligible to participate in the survey to include the D-2 grades.

2. Both price and expenditure surveys were conducted in September/October 2016, although at five European duty stations (London, Madrid, Paris, Rome and Vienna) price data for most of the items in the ICSC basket were sourced from the ECP in accordance with the Commission's decision (ICSC/79/R.10) on the recommendation of ACPAQ. Similar data were collected in New York, the base of the post adjustment system, as presented in document ICSC/ACPAQ/39/R.3. During data analysis, the price data collected in New York were specifically tailored to each location by the ICSC secretariat in order to ensure a fairer comparison. The results of these place-to-place surveys were presented to ACPAQ at its 39th session held in March and June 2017.

3. The Committee noted in March 2017 that both the collection and processing of data from the 2016 place-to-place surveys had been carried out in accordance with the approved methodology, but that certain difficulties[2] had arisen with the use of the ECP price data. The Committee decided to inform the Commission of the new post adjustment indexes for Geneva, Montreal and Washington, D.C. based on the survey results, but it deferred a decision on the remaining five duty stations, pending further study.

4. The ICSC reviewed the ACPAQ report on the survey results at its 84th meeting in March 2017. The Commission approved the recommendations of ACPAQ, although the ICSC secretariat noted at the time that, in the case of Geneva, this would lead to a salary reduction of about 7.5 per cent in United States dollar terms or about 6.7 per cent in Swiss Franc terms, as of the survey date (October 2016).[3] The precise magnitude of the salary reduction could not be determined until the necessary updating of both the survey result and prevailing pay index from the survey date to the survey result implementation date was carried out. Based on the ICSC decision at its 81st session in August 2015 to eliminate the 5% gap closure mechanism, this reduction in pay would be implemented without the transitional measures previously in place.[4] However, the ICSC further decided that the implementation for serving staff should start in August rather than May 2017 and the decrease staggered over three months until November 2017.

5. While the organizations were not formally notified by the ICSC of its decision or of the implementation measures, the information was published on the ICSC website. This resulted in a swift reaction from the Executive Heads of Geneva-based organizations, as well as from staff and staff associations. A letter was sent by the Executive Heads in Geneva[5] on 13 April 2017 to the Chair, ICSC requesting the Commission to provide all relevant information regarding the specific impact that the survey components and the changes to the methodology had on the 2016 survey results. They proposed the deferral of any implementation until such information was available and validated in a process in which representatives of the organizations participated.

6. Subsequently, a meeting was held on 25 April 2017 in Geneva between the Vice Chair of the ICSC and the Executive Heads. Following this meeting, on 28 April 2017, the Executive Heads sent a second letter to the ICSC that reaffirmed their request for deferral and thanked the ICSC for its

[2] The Committee decided to defer recommendations regarding the survey results for London, Madrid, Paris, Rome and Vienna, as part of its plan to assess and account for any systemic differences resulting from the transition from the use of data collected by the ICSC to the use of the European Comparison Programme average prices for those duty stations.
[3] ICSC/84/R.8, para. 100
[4] see paras. 20-27 below
[5] ILO, UNOG, ITU, WIPO, WHO, UPU, IOM, WMO, UNAIDS and UNHCR

commitment to providing the requested detailed information on the specific impact of the changes made to the survey methodology and operational rules governing the post adjustment system prior to the 2016 survey round.

7. The Chair, ICSC replied on 9 May 2017 providing information specifically on the Geneva survey results, as well as providing additional explanation of the results of the 2016 baseline cost-of-living surveys at Headquarters duty stations. In particular, he noted that many organizations and all staff federations had participated in the ACPAQ meetings and were well informed about changes to the methodology and the preparations for the 2016 survey round. Nevertheless, he noted that the organizations might submit additional information at the Commission's 85th session in July in Vienna.

8. Geneva-based organizations considered that the information received from the ICSC was not sufficient to ensure that they could exercise their obligations as established by the principles of international civil service law and evidenced in the jurisprudence of international administrative tribunals before implementing a decision. Acting on the ICSC's commitment to provide the requested data and in the interests of ensuring the results were well understood and areas of specific concern identified, Geneva-based organizations mandated an informal team of three senior statisticians to review the application of the methodology and the data processing work. This team's findings are summarized below in paragraphs 10-30 and are elaborated in greater detail in the full report, which is attached as Annex I.[6] Geneva-based organizations also undertook a review of the legal and managerial issues arising from the survey result. The details of this review are found in paragraphs 31-59 below.

9. During the resumed ACPAQ session in June 2017, similar findings and concerns were identified with regards to other headquarters duty stations (London, Madrid, Paris, Rome and Vienna) and it is expected that the same would apply to the other group 1 duty stations (once those surveys have been completed). It should be further noted that many of the issues identified would also have an impact on group 2 (field) duty stations. The extent of that impact, however, cannot be determined without a case-by-case review.

Review of the Cost-of-Living Survey in Geneva, October 2016

10. Geneva-based organizations requested the informal review team to undertake a targeted review of the ICSC cost-of-living survey in Geneva in order to ascertain whether, from a statistical perspective, the calculations used in the 2016 survey round could be considered of good quality and sufficiently robust to be designated *'fit for purpose'*.[7] This targeted review commenced with a desk review of available documentation supplemented by a mission by three members of the team that was conducted over two and half days between 31 May and 2 June 2017. Thus, it cannot be considered a thorough or comprehensive review of all elements of the ICSC methodology or implementation of that methodology.

[6] Given the opportunity presented by the special session of ACPAQ (29-30 June 2017), a preliminary draft of this report was provided to the ICSC secretariat.

[7] *'Fit for purpose'* is a common statistical term referring to the statistics being of sufficient quality to support how they are used in the context of the user's needs. It is generally recognised that there are several important dimensions to statistical quality beyond accuracy, such as relevance; credibility; timeliness; accessibility; interpretability; coherence and cost-efficiency. For example, see the *Quality Framework and Guidelines for OECD Statistical Activities (2011)*, para. 8. (http://www.oecd.org/std/21687665.pdf).

11. The findings below should not be considered as reflecting a statistical validation in pursuit of perfection, but rather as a due diligence effort conducted by the organizations (in cooperation with the ICSC) in order to ensure legal "defendability" of results given the impact on staff. The current system has many strong features which should be retained and in some cases enhanced to further strengthen the credibility of the statistics. With that said, the informal review team has identified a range of issues that are believed to require immediate correction as they represent errors or statistically invalid application of the approved methodology that have statistically biased the results of the 2016 round of cost-of-living calculations for Geneva. Beyond these, there are other issues which require attention for which the team has provided comments which can hopefully be a useful input to the ICSC secretariat for future review activities. A more detailed discussion of the various issues highlighted is presented in the full report of the informal review team presented at Annex I.

12. All estimates of impact are approximate and can only be confirmed through full recalculations undertaken by the ICSC. The issues identified are in many cases also of direct relevance to other duty stations, but the degree of impact would need to be assessed on a case-by-case basis.

Issues requiring immediate correction:

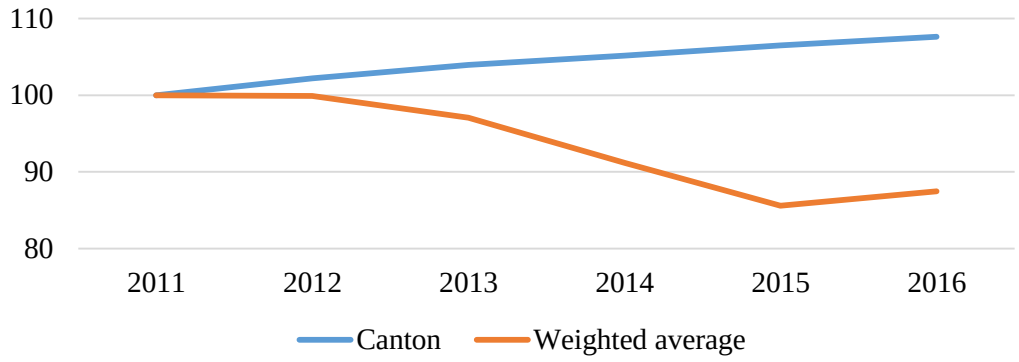
Rents:

13. By far the most important element of any immediate review of cost-of-living calculations is the rental component. With a weight of almost 23%, rents are by a distance the single largest basic heading, meaning that any movement in this sub-index will contribute close to a quarter of the overall change in the PAI for Geneva. Two issues regarding the rents calculations are considered to require immediate attention, as discussed further in section 2.1 of the report at Annex I:

a. The aggregation formulae used by the ICSC in the calculation of the housing sub-index are well-defined and established, namely: Laspeyres, Paasche and Fisher being the geometric mean of the Laspeyres and Paasche indexes. These formulae require that **expenditure** weights (i.e. the amount of spending on each type of property) be used to reflect the relative importance within each index of the different types of property rented by UN staff. The ICSC secretariat confirmed, however, that **quantity** weights (i.e. the number of staff living in each type of property) were used. Correction of this error would increase the Geneva PAI by 1.3%.

b. The rent data provided to the ICSC by the ISRP has generated implausible results for Geneva. Over the period 2011 to 2016, this indicated a drop of approximately 12% in average rents for Geneva. Official data, which are considered very robust, showed an increase of close to 8% over the period (see figure 1). Detailed examination of this issue does not indicate that there is any methodological reason or other explanation which would justify such a difference in trend. On this basis, the review team concluded that its use is not suitable, has clearly created a substantial downward statistical bias in the Geneva PAI and should be corrected. The estimated impact of using more statistically robust data is an increase in the Geneva PAI of approximately 2.5%.

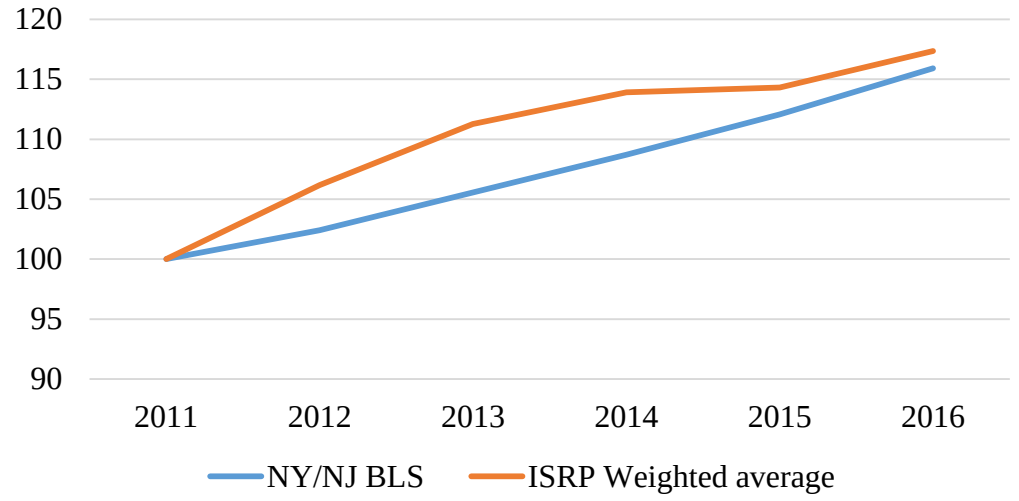
Figure 1: Geneva rents trend
Canton of Geneva rents index (all rentals) compared with ISRP rents data (weighted average)
2011 to 2016 (2011=100)



Source: Statistical Office of Canton de Genève and calculated trend based on data in ACPAQ/39/R.4

14. By contrast the ISRP data for New York is quite consistent with trends in official data from the Bureau of Labour Statistics (BLS). In fact, it has shown a slightly higher increase between 2011 and 2016 than the BLS series despite indications that some 'cheaper' areas had been included by the ISRP (see figure 2). This suggests that difficulties with ISRP data are not necessarily universal and in the case of the Geneva to New York comparison have only impacted Geneva negatively.

Figure 2: New York rents trend
BLS (CPI) rent index compared with ISRP rent index 2011 - 2016 (2011=100)



Source: BLS (NY/NJ Rental Index) and authors own calculations based on data in ACPAQ/39/R.3

15. If corrected, the combined impact of these two issues would be to increase the Geneva PAI by approximately 4%, bringing the remaining decrease in the Geneva pay index to 4% or less, well below the 5% threshold for a change in the pay index.

Domestic services:

16. A statistically biased editing process was used by the secretariat which has incorrectly lowered the index for Geneva. This is described in more detail in section 2.2 of the report at Annex I. Broadly speaking, this involved removing all values above a threshold of CHF 60 per hour for part-time domestic services in the case of Geneva, but leaving in high values for New York (for example $240 an hour). The review team does not consider that there is any valid statistical reason to adopt such an imbalanced approach. Different methods are available to correct this. In addition, an invalid aggregation method has been used, as the ICSC applied quantity weights to aggregate the lower level indexes to the overall index for domestic services. This approach is inappropriate for aggregation of the indexes between part-time and full time, as they represent very different levels of expenditure. The combined estimated impact of correcting these issues would be to increase the Geneva PAI by between 0.4% and 0.5%.

Education:

17. A data entry error was discovered during the mission at the ICSC. The impact of this was low (less than 0.1% of the Geneva PAI) but should nonetheless be corrected. However, the review team found in discussion with the ICSC that the correction of errors once a calculation is 'locked' is very difficult and that the attempt to do so would have uncertain consequences given the structure of the systems and processes in use. In other words, it is difficult to undo the error in any real sense and salaries would continue to be adjusted, knowing that the statistics informing that decision were incorrect. The system could be strengthened by establishing clear practices and a revision policy for cases where errors are discovered or other revisions become necessary, something which can arise in statistics.

Other important issues requiring review:

18. In the detailed report, the review team highlights a range of additional issues which should be addressed urgently. However, immediate corrections cannot be made with available data and it is not possible to estimate their impact. They include:

a. An inconsistent mix of aggregation formulae is used, some of which are inappropriate for the purposes in question and can have major impact on results generated. This relates to formulae used at both the lower and higher levels of aggregation and goes beyond the specific issues highlighted for rents and domestic services. These are discussed in the report in Annex I and in more technical detail in the appendices to that report, and alternative approaches are suggested.

b. Use of low numbers of items, relative to similar exercises undertaken by the World Bank and Eurostat, and too few price observations create weak averages (for example only one used price observation for various items). This in turn leads to unclear representativeness of the results and can be clearly linked to excessive and undesirable volatility. This can also be considered to be related to use of suboptimal aggregation methodologies. The volatility observed appears to have created a low statistically biased result for Geneva for the 2016 round which cannot be explained by approved changes in methodology. This issue is discussed in detail in section 3.1 of the report at Annex I. The

scale of impact cannot be estimated with available data. In the absence of updates to the methodologies and practices, the review team considers it likely that continued excessive volatility will be observed with other duty stations and for future rounds. This would cause inappropriate adjustments of salary not truly justified by the statistics and would in turn yield unpredictable results.

19. There are clear difficulties with the approach to comparing costs/prices between Geneva and New York for a number of important items including Education and Medical Services among others (see sections 3.2 to 3.5 of the report at Annex I). The chosen approaches do not provide a comparison which is representative of staff expenditures and should thus be updated following a review. Information is available which would allow more appropriate comparisons to be made. Given available information, it is considered that more appropriate approaches would increase the index for these items for Geneva.

Gap closure measure:

20. In 2015, the ICSC agreed upon changes to the existing gap closure measure. Specifically, it was agreed to remove the 5% 'cushion' whereby in case of decrease in the pay index of more than 5%, 5% would be added back to reduce the impact on staff.

21. One of the arguments put forward for the removal of the measure was to avoid distortions due to non-application of the actual results of surveys in full. However, the current system introduces other clear distortions which should be noted, such as the possibility that a duty station with a higher theoretical cost of living ends up with lower pay; an example is presented in the detailed report to illustrate this.

22. With that said, we must consider why there is a need for a gap closure measure. A high level of quality is needed for statistics used for sensitive purposes, such as pay adjustment. However, all statistics will be subject to a margin of error. This is evidently the case for spatial price comparisons such as the cost-of-living surveys of the ICSC.

23. Beyond the clear errors highlighted by the review team, there are many areas of uncertainty which raise questions about the reliability of the statistics generated. This includes cases where statistically biased results appear to have been generated plus elements of methodology and practices which create undesirable and unjustifiable volatility in estimated price comparisons. It is the review team's view that with respect to the 2016 round, this statistical bias has, on balance, worked against Geneva. In fact, it is considered possible that the full calculated decrease in the pay index could be a result of statistically biased results generated from multiple elements of the calculation.

24. However, in the presence of some weak methodologies and practices, statistical bias could work in different directions over time creating a relatively high level of ongoing uncertainty. The gap closure measure is one feasible means to avoid a significant negative impact on staff from results which have a margin of error that cannot easily be estimated.

25. This uncertainty and margin of error make the reinstatement of gap closure a high priority in order to allow for the fact that, regardless of any improvements made, a range of error will continue to exist and does not offer a stable base upon which to cut staff pay in a very precise manner.

26. Another way to express this is that the current system would see a 5.1% pay decrease fully implemented while a 4.9% decrease would lead to no change. It will never be the case that the statistics generated from the cost-of-living surveys can be sufficiently precise to justify this difference in treatment. In fact, it could be argued that the margin of error with current methodologies and processes could be far in excess of 5%, however, in recognition of the precedent of a 5% gap closure measure, this is a reasonable approach to protect staff from the worst effects of full implementation of uncertain results.

27. In summary, it is clear that a high standard of quality is required of statistics used to adjust pay. The removal of the gap closure measure in 2015 made it effectively impossible for the statistics generated to meet the level of quality required to justify their usage for the purpose in question. Reinstating the gap closure measure creates an allowable margin of error which can be achieved if the right methodologies and processes are applied.

Summary:

28. The review team has identified issues of differing impact and source. Certain issues are believed to require immediate correction as they make the current results clearly and demonstrably statistically biased and improved alternatives can be calculated. The correction of these issues is estimated to increase the Geneva PAI by over 4%. This would reduce the remaining reduction in the pay index as of May 2017 to less than 4%. In this event, there would be no change in the Geneva pay index as it is below the 5% threshold for change.

29. A variety of other issues are identified, which should be reviewed and on balance are considered to have generated a low statistically biased result for Geneva on this occasion.

30. Beyond the details of the statistical measurement, the review team would recommend a variety of improvements in practices including the range of documentation published and development of clear policies on revisions among others. It can be noted that none of the current review mechanisms, while an important element of the current system which should be continued, could have identified the issues highlighted as requiring immediate correction as available documentation did not facilitate this.

Legal and Managerial Implications

Discretion in implementing ICSC decisions and recommendations

31. As members of the UN common system, the agencies represented in the Human Resources Network have accepted the mandate of the ICSC and, accordingly, they implement ICSC decisions and recommendations following each organization's own constitutional obligations and applicable procedures. The organizations' commitment to implement the ICSC decisions and recommendations does not necessarily result in automatic implementation as legal and/or managerial considerations could require or support a decision not to implement the decision/recommendation of the ICSC, either in whole or in part

32. According to the case law of the International Labour Organization Administrative Tribunal (ILOAT)[8], the organizations have a duty to analyse the ICSC decisions and recommendations and to give effect to them only if satisfied that they are lawful and that doing so would not impair the rights of staff (judgment 1713, consideration 3; judgment 1265, consideration 24; judgment 825, consideration 18). Accordingly, while the UN organizations that have recognized ILOAT's jurisdiction will align the conditions of service of their staff as far as possible and practicable with those of other UN organizations, they also have an obligation to ensure that these conditions are lawful.

33. Therefore, before taking any measure to implement the ICSC decision concerning the post adjustment multiplier, the organizations have to ensure that such a decision is lawful and, as applicable, consistent with the principles of responsible management and fair employment practices. For the reasons explained below, the organizations are not in a position to come to a conclusion, in light of the information currently available, that the decision to modify the post adjustment multiplier meets the technical, legal and managerial requisites.

34. While the unlawfulness of an ICSC decision/recommendation would be a sufficient reason *per se* for not implementing the proposed change, the legal framework of some Geneva-based organizations also expressly provides for a certain degree of discretion on whether to implement a decision/recommendation and on how/when to implement it.

Legal considerations

35. The ILOAT has developed consistent case law regarding the review of the legality of administrative decisions taken on the basis of ICSC decisions/recommendations. Notably, in judgment 1821, consideration 7, it ruled as follows:

> *The principles governing the limits on the discretion of international organisations to set adjustments in staff pay have been well established in a number of judgments. Those principles may be concisely stated as follows:*
> *(a) An international organisation is free to choose a methodology, system or standard of reference for determining salary adjustments for its staff provided that it meets all other principles of international civil service law: Judgment 1682 (in re Argos and others) in 6.*
> *(b)* ***The chosen methodology must ensure that the results are "stable, foreseeable and clearly understood****": Judgments 1265 (in re Berlioz and others) in 27 and 1419 (in re Meylan and others) in 30.*
> *(c) Where the methodology refers to an external standard but grants discretion to the governing body to depart from that standard, the organisation has a duty to state proper reasons for such departure: Judgment 1682, again in 6.*
> *(d) While the necessity of saving money may be one valid factor to be considered in adjusting salaries provided the method adopted is objective, stable and foreseeable (Judgment 1329 (in re Ball and Borghini) in 21), the mere desire to save money at the staff's expense is not by itself a valid reason for departing from an established standard of reference: Judgments 1682 in 7 and 990 (in re Cuvillier No. 3) in 6."* (Emphasis added)

[8] Some Geneva-based organizations have recognized the jurisdiction of the ILOAT to adjudicate on complaints filed by staff members against administrative decisions and are accordingly subject to its case law.

36. The requirements of lawfulness of the ICSC decision on the post adjustment multiplier include ensuring that the methodology used is void of any flaw, is correctly applied and is compliant with the principles of stability, foreseeability and transparency. In addition, as emphasized by the Tribunal in Judgment 1713 (consideration 3), when an organization gets notice of a recommendation of the ICSC, it must ensure that the implementation of the recommendation would not impair the rights of its staff:

> *As the FAO acknowledges in its reply, "when it gets notice of the Commission's recommendations on salary scales it has a duty to analyse them and to give effect to them only if satisfied that they are lawful and that doing so would not impair the rights of its staff".*

37. It is also important to bear in mind that case law requires all administrative decisions to be properly motivated, which requires organizations to verify that the ICSC's explanations are sufficient for that purpose.

Ensuring that the methodology is flawless and correctly applied

38. In judgment 1713, the Tribunal also expressly recognized that while the ICSC must be allowed some discretion over the method in choosing figures of local pay, the ILOAT will still review the exercise of it. It stated that *"*[t]*he decision impugned may not stand if, say, it overlooks or misconstrues some particular factor, or if some method is applied for the wilful contrivance of lower figures of local pay, or if corners are cut for the sake of saving time, but to the detriment of staff interests."* (consideration 8).

39. In judgment 1765, the Tribunal found that the ICSC made a mistake in reckoning the multiplier used to work out the post adjustment in Geneva. The Tribunal observed the following (consideration 8):

> *[The Organization] has the duty of checking the lawfulness of any decision by another body on which it bases its own decision.* ***So too must it check the adequacy of action by that other body to correct any mistake it may have made, and make sure that such corrective action respects the rights of staff.*** (emphasis added)

40. Before implementing the ICSC decision regarding the revised post adjustment multiplier for Geneva, the organizations, therefore, have the duty to ensure that the methodology was flawless and correctly applied.

41. As emphasized above in paragraphs 10-30 and in Annex I to the present document, the informal review team has identified errors or statistically invalid application of the approved methodology that have statistically biased the results of the 2016 round of cost-of-living calculations for Geneva. The informal review team also singled out a number of methodological changes introduced since 2010 which have increased the instability and volatility of the indexes making up the cost-of-living comparisons. According to the informal review team, the identified methodological problems compromise the quality and robustness of the overall ICSC calculations.

42. These results illustrate that the lawfulness of the revised post adjustment multiplier could be challenged on the basis of the methodology used and the errors encountered in its application. In any case, the identified errors should be investigated and corrected.

Ensuring that the methodology respects the principles of stability, foreseeability and transparency

Stability

43. In the case at hand, the methodology for conducting place-to-place surveys for group 1 duty stations (which includes all headquarters duty stations) has suffered changes which resulted, *inter alia*, in the removal of several stabilizing factors used in the previous methodology to establish post adjustment multipliers. The rationale for these changes and for the removal of the stabilizing factors remains ambiguous as the conditions in the Geneva duty station have remained stable and, therefore, challenges would be posed to the organizations in defending the lawfulness of the revised methodology.

44. Taking into account the magnitude of the proposed decrease in the post adjustment multiplier, one could successfully argue that the changes applied to the methodology did not ensure stable results – and this regardless of whether the methodology was correctly applied or not. A valid methodology should not produce results that vary significantly when conditions have remained stable.

Foreseeability

45. No Geneva-based organizations anticipated a decrease of the post adjustment multiplier of this importance. This is why it was decided to address on 13 April 2017 a common letter to the ICSC to express the organizations' deep concerns about the proposed decrease, emphasizing "*the exceptional magnitude of the impact on the conditions of employment of* [their] *staff"*. These concerns were furthermore expressed during the meeting with the Vice-Chairman of the ICSC and the Chief of the Cost-of-Living Division on 24 and 25 April 2017 and in the subsequent letter to the ICSC dated 28 April 2017.

46. Moreover, to date the organizations do not know or have even just an indication of the impact the revised methodology would have on other duty stations, in particular field duty stations. Such unpredictability is not acceptable and goes against the principles of international civil service law as well as those of responsible management. The organizations are very concerned about some of the developments concerning post adjustment multipliers in several field duty stations during the past five years and consider that there is a need to carefully review the respective surveys, the way the methodology was applied and the results in order to establish whether corrective action would be required.

Transparency

47. A good faith and meaningful consultation on the changes applied to the methodology and the preparations for the 2016 survey round, as referred to in the ICSC letter of 9 May 2017, would have had to include a presentation of scenarios demonstrating the impact of each change and a constructive dialogue on the proposed changes. Instead, numerous piece-meal changes were made to the methodology over several years without clear rationale; these changes included the removal of stabilizing factors, such as the removal of the 5% gap closure measure while maintaining the 5% buffer zone; the organizations' only involvement was as observers and once the survey was finished no indication of the results and implications was given. When figures were finally provided to the organizations' representatives at the ACPAQ session in February 2017, these figures did not provide the actual impact on salaries and at no point were the Executive Heads of Geneva-based organizations informed by the ICSC, in spite of the

obligation of the ICSC to consult with the Executive Heads of the concerned organizations, prior to a determination being made, as expressly provided by the ICSC Rule of Procedure 33.[9]

***Ensuring that the revised post adjustment multiplier respects the acquired rights of the staff members*[10]**

48. The basic principle for determining an acquired right is that an amendment to an official's detriment of a provision governing her or his status constitutes a breach of an acquired right only if it adversely affects the balance of contractual obligations by altering fundamental terms of employment in consideration of which the official accepted an appointment, or which subsequently induced her or him to stay on (ILOAT judgment 3676, consideration 14).

49. In judgment 832, consideration 14, the Tribunal identified three tests to determine whether the altered term is fundamental and essential:

> *The first is the nature of the altered term. It may be in the contract or in the Staff Regulations or Staff Rules or in a decision, and whereas the contract or a decision may give rise to acquired rights the regulations and rules do not necessarily do so.*
>
> *The second test is the **reason for the change**. It is material that the terms of appointment may often have to be adapted to circumstances, and there will ordinarily be no acquired right when a rule or a clause depends on variables such as the cost-of-living index or the value of the currency. Nor can the finances of the body that applies the terms of appointment be discounted.*
>
> *The third test is the **consequence of allowing or disallowing an acquired right**. What effect will the change have on staff pay and benefits? And how do those who plead an acquired right fare as against others?* (emphasis added)

50. In judgment 3571, after having recalled its well-established case law on acquired rights, the ILOAT found that *"**on account of its magnitude**, the alteration of the basis for calculating the complainant's contributions to the Eurocontrol Pension Scheme breached a fundamental and essential term of employment"* (emphasis added).

51. Also, in judgment 3623, where it was found that there was no breach of acquired rights, the ILOAT judged that *"[t]he changes in circumstances which may require the rule to be amended must be reasonable and the changes have to balance the interests of the employees and the Organisation. The interest of current and future employees who are not currently affected by the rule but shall be in the future is also to be taken into account by the Organisation."*

---

[9] ICSC Rule of Procedure 33: "[…] *Neither the Commission nor any member to whom a function has been delegated shall make a substantive determination, other than a routine revision of daily subsistence allowance rates or of the classification of duty stations for the purpose of applying post adjustments, before having sought the views of the executive heads of the participating organizations concerned on the financial and administrative implications of implementing that determination and having received a report of the Executive Secretary thereon* […]"

[10] ICSC Statute, Article 26: *"The Commission, in making its decisions and recommendations, and the executive heads, in applying them, shall do so without prejudice to the acquired rights of the staff under the staff regulations of the organizations concerned."*

52. Based on the case law, including the judgments quoted above, it is reasonable to assume that there are high risks that the ILOAT, should a complaint be brought before it (which is very likely if the revised post adjustment multiplier is implemented), would recognize that the implementation of the ICSC decision does indeed constitute a breach of the acquired rights of the staff members, despite the fact that the post adjustment multiplier is, by definition, subject to change:

a. As it could prove to be difficult to defend why the changes made to the methodology are well-justified (especially taking into account that there has been no appreciable reduction in local costs); and/or
b. On account of the magnitude of the change and the severe financial implications for the staff members. The proposed cut would also add to the reduction in the overall income resulting from elements of the new compensation package for staff in the Professional and higher categories as implemented in 2017.

53. Geneva-based organizations consider that there is significant legal risk that a decision to implement the ICSC decision would be found by the ILOAT unlawful and in breach of the staff members' rights, if such a decision were to be challenged before this instance; and, as already mentioned above, it is very likely that staff members would indeed contest such a measure (including a considerable risk of class appeal).

Managerial considerations

54. Geneva-based organizations are also deeply concerned about how the implementation of the ICSC decision would contravene the principle of sound, responsible and good administration of its most essential resources. In addition, they have noted that the transitional measures decided by the ICSC for the implementation of its decision[11] are not in line with previous decisions made by the ICSC in similar situations, such as:

a. ICSC report of 1988, para. 40, 41 (c)

The Commission noted that in the case of three of the duty stations listed above, namely London, Paris and Rome, the RCF was currently applicable. The Commission further noted that there were a number of field duty stations where surveys had been conducted and where the RCF was currently applicable. It was recognized that, if the RCF were eliminated at these duty stations in conjunction with the implementation of the survey results, staff members would experience losses in take-home pay in local currency terms. In general, bearing in mind the question of staff-management relations, the traditional policy of the Commission had been not to reduce staff take-home pay as a result of cost-of-living surveys, but to freeze it until cost-of-living movements caught up with the level of take-home pay.

In view of the foregoing, the Commission decided that in the case of duty stations where the RCF and floor protection measures applied, the following procedures should be used when implementing the results of new place-to-place surveys: (c) Where the post adjustment multiplier resulting from the procedure described in subparagraph (a) above was lower than the existing

[11] A first reduction of 5% on 1 August 2017 and a second reduction of 2.9% on the 1 November 2017, resulting in a cumulative change in net pay of 7.7%

multiplier, take-home pay expressed in local currency should be frozen until a new full class of post adjustment was reached, using the exchange rate on the date of implementation of the survey results.

b. ICSC report of 1980, para. 146

The Commission has continued to perform the function entrusted to it under article 11 (c) of its statute, of establishing the "classification of duty stations for the purpose of applying post adjustments". After extensive investigation under the aegis of the Commission's Advisory Committee on Post Adjustment Questions (ACPAQ), the Commission decided to reduce the post adjustment index for Geneva by one class from its October 1979 level, which was 244.1 at the exchange rate of SwF 1.69 to the US dollar. The new index for October 1979 would then become 232.5 at the same exchange rate. The old index would remain frozen until the new index reached 244.1. Thereafter, customary updating procedures would recommence. Adjustment for currency fluctuations would continue without interruption, as heretofore.

Negative effects on staff morale

55. Implementation of the revised post adjustment multiplier would have a significant impact on staff morale, motivation and productivity, which is clearly not in the organizations' best interests.

56. Since the announcement of the ICSC decision, staff members have expressed their concerns and clear dissatisfaction on several occasions. A number of meetings and rallies of staff members were organized by staff associations of different organizations and a work stoppage took place on 16 June 2017. A survey initiated by the WHO headquarters Staff Committee indicated that over 90% of the respondents were in disagreement with the ICSC decision.

Adverse impact on the employment competitiveness of the organizations

57. The demands on organizations and staff are steadily increasing. To be credible hubs of innovation and efficiency and to deliver complex mandates, including the implementation of Sustainable Development Goals, requires that organizations attract and retain talent sourced on global employment markets. Even without a reduction in post adjustment, this is increasingly challenging in Geneva as the UN common system compensation package is losing its competitiveness. This is particularly the case in key professional areas, for example in information technology and translation services. In addition, for certain highly specialized professional skills needed by organizations of the UN common system, for example in the area of economics and law, other international organizations and the private sector offer more attractive employment terms.

Impact on the difference of remuneration between the General Service and the Professional categories

58. The implementation of the revised post adjustment multiplier would also give rise to serious concerns about how this decrease in remuneration for Professional staff, on top of changes to the compensation package in 2017, would further exacerbate the convergence of Professional and General Service remuneration levels in Geneva, in direct conflict with the demands and levels of responsibility that distinguish these two categories of staff.

Guidance from Governing Bodies

59. Depending on the outcome of the Commission's consideration of the issues raised in this paper, the organizations may need to provide detailed explanations of the situation to their respective governing bodies for confirmation on management actions taken and for guidance on future actions.

Conclusions

60. In light of the significant technical, legal and managerial risks foreseen in the implementation of the revised post adjustment multiplier, Geneva-based organizations are strongly of the view that the revised methodology must be reviewed and assessed in line with the recommendations made by the informal review team. Based on such a review, changes and corrections should be made to the current round of results. As needed, existing methodologies should be reviewed and new methodologies should be developed, ensuring that, in the future, the methodology meets the requirements of lawfulness and respect of staff members' rights, regardless of their duty station, and takes full account of the managerial considerations identified above in addition to the statistical quality requirements essential to ensure the statistics are fit for purpose.

61. In the meantime, the best way to move forward appears to be either to (1) make corrections as proposed by the informal review team and reinstate the 5% augmentation in the gap closure measure, thus re-introducing a stabilizing factor to provide for a reasonable margin of error, (2) freeze the April 2017 post adjustment multiplier pending a more thorough review, or (3) continue to apply interim adjustments following the method applied for these since the previous place-to-place survey (2011), until the methodology has been corrected or a new methodology developed and the post adjustment multiplier recalculated accordingly or, if the 2016 methodology and survey results are validated in the review, until the post adjustment multiplier catches up with the April 2017 multiplier.

62. Given the concerns with regard to the statistical accuracy of the methodology and its application to the 2016 survey round and the lawfulness and resulting legal risks of any actions based thereon, Geneva-based organizations are not in a position to consider a staggered implementation of a post adjustment multiplier resulting from a process that currently does not meet the organizations' legal or managerial requirements.

63. In accordance with these conclusions, Geneva-based organizations submit that it is incumbent upon the ICSC to:

a. Consider seriously the conclusions reached by the informal review team as developed in Annex I of the present document and take immediate appropriate actions to correct the identified errors;

b. Suspend its decision to implement the survey result for Geneva as of 1 May 2017 with a staggered approach over several months;

c. Review the survey methodology in the light of the conclusions and recommendations presented in this paper in close consultation with the organizations of the UN common system;

d. Rescind its decision to modify the gap closure measure in determining the post adjustment multiplier applicable to a duty station by abolishing the 5 per cent augmentation of the post adjustment index derived from negative place-to-place survey results when the difference between the existing pay index and the new post adjustment index resulting from the survey is more than 5 per cent; and

e. Notwithstanding (d) above, review in consultation with the organizations the issue of transitional measures and other operational rules of the post adjustment system, as may be required.

64. Geneva-based organizations believe that once these actions have been taken, they will then and only then be in a legally-defensible position to implement the ICSC decision on the Geneva survey result.

ANNEX I

REVIEW OF THE COST-OF-LIVING SURVEY IN GENEVA, OCTOBER 2016

Report of Geneva-based organizations Informal Review Team

June 2017

Kieran Walsh, ILO

Steve MacFeely, UNCTAD

Mick Silver, International Consultant

Valentina Stoevska, ILO

Report of Geneva-based organizations Informal Review Team

June 2017

Acknowledgements

1. The review team would like to thank the International Civil Service Commission (ICSC) for assisting us in a review of the methodology ad providing it with access to the microdata. We would like to acknowledge the cooperation and transparency we received. In particular, we would like to thank Mr. Ibrahim Yansaneh, Mr. Roberto Pagan, Ms. Cleopatra Todis, Mr. Kari Manninen and Mr. Ignacio Vera for their time, expertise and openness in assisting us with this review. This review recognises the efforts and professionalism of ICSC staff in undertaking what is a very complex and difficult task.

65. The team would also like to thank the management of the ILO and UNCTAD, for the financial support facilitating a mission to New York.

66. The review team would also like to acknowledge the considerable statistical assistance provided by Ms. Yumiko Mochizuki (UNCTAD), Mr. Bojan Nastav (UNCTAD), Ms. Nour Barnat (UNCTAD) and Mr. Liam Hanley (UNCTAD).

Confidentiality

67. In line with best statistical practice, the authors of this report have signed a confidentiality agreement ensuring that no individual data will be disseminated. Consequently, only aggregate data and analyses are provided in this report.

Terms of Reference

68. The Geneva Human Resources Group requested the informal review team to undertake a targeted review of the ICSC cost-of-living survey in Geneva in order to ascertain whether, from a statistical perspective, the calculations used in the 2016 survey round could be considered of good quality and sufficiently robust to be designated ‘fit for purpose’.

69. This targeted review commenced with a desk review of available documentation supplemented by a mission by three members of the team which was conducted over two and half days between 31 May and 2 June 2017. Given the relatively short mission time and its motivation the focus was by necessity on basic headings of the PAI that had a particularly negative impact on Geneva. This review cannot thus be considered a thorough or comprehensive review of all the elements of the ICSC methodology or implementation of that methodology. The focus on these basic headings should not be taken as an affirmation of the methodology used generally or with regard to other basic headings. On the contrary, given the concerns raised in this report, an implicit question mark must be raised regarding the methodology and its’ application for other basic headings and procedures. Consequently, the ICSC would benefit from a comprehensive review of the PAI methodology aimed at confirming that the methodology for other basic headings are compiled in line with international standards and best practice. There were some aspects of the methodology that required clarification and we are grateful for the assistance provided by ICSC in this regard. However, the need for clarification on what turned out to have a

substantive bearing on salient issues, begs a review of areas not fully studied in this report. Such a review would further serve to recognize the sound footing on which much of the ICSC's work is based and further enhance its credibility.

70. The scope of this report centres on relativities between Geneva and New York. However, several of the problems identified are systemic in nature and will clearly have implications beyond those two duty stations. Unfortunately, without access to the microdata for each station we cannot assess in any detail the impact for duty stations beyond Geneva.

Accuracy of our calculations

71. A number of alternative calculations are presented in this report. While the authors are confident of the accuracy of those estimates and their impact on the overall PAI is correct (bearing in mind the assumptions made), those calculations should be first be tested within the ICSC system, to ensure they are precise.

'Approved' ICSC Methodology

72. The approved ICSC methodology is described in 'The Post Adjustment System - what it is, how it works' (ICSC, 2011), albeit it with some updates approved by ACPAQ for the 2016 round. This is very important, as the ICSC have stressed that their role is to implement the 'approved' methodology. Consequently, in their view, only problems with regard to implementation of the approved methodology are within scope for discussion. In other words, methodological issues, even if problematic, do not invalidate any given result. Ostensibly this does not seem unreasonable as there must be an agreed approach to making cost-of-living comparisons. A problem exists however, as the approved methodology described is not consistent with what is implemented in practice in a number of cases. Furthermore, some descriptions are technically inaccurate. The most striking example is the reference to the use of a modified Walsh index in the methodological description (ICSC, 2011, p.21). Nowhere in the calculations is a Walsh (or modified Walsh) index used - the ICSC confirmed this during our mission to New York (see Appendix 3). Yet the ICSC stated in their implementation report that "In accordance with previous ACPAQ recommendations and corresponding Commission decisions, a RTPC - adjusted modified Walsh formula was used in the calculation of the cost-of-living indices for all headquarters duty stations and Washington D.C." (ICSC 2017, P.39). Similar inaccurate references can be found with regard to 'common weights'. Such inconsistency undermines any arguments that an 'approved' methodology has been implemented while also making assessment of application of methodologies more difficult.

Quality

73. Most official statistics are compiled under the auspices of a specified code of practice and typically guided by a quality framework. These codes and frameworks outline the standards expected that allow estimates to be described as *'fit for purpose'*. Depending on the purpose of the statistic, the standards set may be more or less onerous. Considering the importance of the work of the ICSC, it is surprising that no explicit quality dimensions and standards have been agreed and published. Furthermore, given the sensitivity and very direct application of the results the quality standard targeted for ICSC statistics must be at a high level.

74. While there is no single agreed definition of statistical quality nor any internationally agreed standard set of quality dimensions, there is a high degree of agreement across the quality frameworks and standards published by most international organisations, including the United Nations *National Quality Assurance Framework* (2012), the OECD's *Quality Framework and Guidelines* (2011) and Eurostat's *European Statistics Code of Practice* (European Commission, 2011).

75. In general, these frameworks categorise dimensions of quality into three groups: output quality; process quality; and institutional quality. Output quality includes dimensions such as: relevance; accuracy; reliability; coherence; timeliness; punctuality; accessibility; and interpret-ability. Process quality includes dimensions such as: sound methods and systems and cost efficiency. Institutional quality includes dimensions such as: objectivity; impartiality; transparency and credibility.

76. More particularly, there are extensive internationally-agreed Manuals of standards on price index number measurement including each of consumer, producer, export and import price indices. The International Labour Office, along with Eurostat, the International Monetary Fund (IMF), Organisation for Economic Co-operation and Development (OECD), United Nations Economic Commission for Europe (UNECE), and the World Bank are co-signatories of such Manuals. Indeed the 2004 *Consumer Price Index Number Manual* (ILO, *et al*., 2004) while written under the aegis of the United Nations' Inter-secretariat Working Group on Price Statistics (IWGPS), however, was led and published by the ILO. These Manuals were concerned with price change measurement over time. Manuals and Handbooks for the World Bank's International Comparison Program (ICP) are available for spatial measures of price comparisons, such as that undertaken by the ICSC. While the task of the ICSC has its own particular challenges, there is no absence of sound internationally-agreed understanding of the principles and practice of price comparisons from which the review team could draw.

77. The issues raised in this review are, by the nature of the task, technical but are based on sound and well-established methodological standards as to what is and what is not right for proper measurement. Moreover, an erroneous application of these standards makes a difference.

Introduction

78. Following the 2016 round of price and family expenditure surveys, the ICSC at their eighty-fourth session announced that salaries of P and D grades based in Geneva were to be reduced by 7.7%. Unsurprisingly this caused some shock as there was little anecdotal evidence to support a change of this magnitude.

79. Preliminary reviews of the published ICSC results revealed a number of results that appeared inconsistent with other publically available official statistics and economic indicators. As a consequence, questions arose regarding the quality of the ICSC calculations and the wisdom of the subsequent decisions based on those data.

ANNEX I

80. Between 31 May and 2 June 2017, the authors of this report visited the ICSC offices in New York in order to better understand the methodology used, clarify some immediate matters of concern and review some of the individual data and calculations used to compile the ICSC report12. In the short time available, only a limited amount of methodological issues and data could be reviewed. Nevertheless, a number of issues were identified that in our opinion call into question the validity of the overall result. These issues are outlined in this report.

81. The report is divided into five Sections. Section 1 of the report provides a summary of the most important key messages emerging from our investigation. Section 2 identifies, in more detail, issues that we argue must be addressed immediately if the ICSC calculations are to be considered 'fit for purpose', as they concern implementation of the 'approved' methodology. In our view, there is absolutely no reason why these changes cannot be quickly made, as our estimates will demonstrate. Section 3 highlights a number of methodological issues that are sufficiently problematic to be considered urgent and must be addressed before the next round of the place-to-place surveys. The ICSC have made clear, that from their perspective, only problems regarding the implementation of the 'approved' methodology were within scope for discussion or amendment. While ostensibly a reasonable position, it not always easy to ascertain what the approved methodology is exactly, as parts of the published documentation is clearly out-of-date begging the question what else may not be up-to-date. Furthermore, several worrying methodological problems have been identified that in our opinion compromise the quality and robustness of the overall ICSC calculations. Section 4 provides a brief discussion of the 'gap closure measure' which was changed in 2015. The fifth and final section provides a summary conclusion and makes some general recommendations regarding the ICSC programme and procedures. Three technical appendices are also attached for readers who may wish a formal exposition of the issues highlighted elsewhere in the report.

[12] ICSC/ACPAQ/39/R.4

Section 1 - Key Messages

Conclusions

- The authors have identified a number of serious calculation and systemic errors in the compilation of the ICSC results. Consequently, the ICSC calculations as currently compiled for Geneva cannot be considered of sufficiently good quality to designate them *'fit for purpose'*.

- The implementation by the ICSC does not always correspond with the 'approved' methodology described in the formal documentation. Furthermore, despite a large volume of documentation being available, many important compilation methodologies are not described anywhere in the formal documentation.

- A number of methodological changes introduced since 2010 have increased the instability and volatility of the indices used to calculate the cost-of-living comparisons. In the 2016 round, these changes appear to have almost universally reduced the Geneva PAI.

Specific issues in measurement

- The index number formula used at the higher (weighted) level are, in two important cases wrongly formulated due to their use of quantity, as opposed to expenditure, weights (Rent and Domestic Services). In both cases the errors arising are significant and, moreover, erratic.

- A variety of index number formulas are used for different basic headings. There is no rationale given for the choice between formulas and no immediately apparent reason, for example, why geometric weighted aggregates are used in one instance and arithmetic weighted ones in another, or why Geneva or NY weights are used, or some formulation involving a symmetric mean of the two city's weights. These choices have important impacts on the overall results. While this report will focus on rental and domestic services, we believe the problem is endemic.

- The problem is not just with the plethora of formula used and inappropriate weights, but that in important instances the formula used cannot be justified. Indeed, in some substantive instances – for "domestic services" and aggregation of the five major components at the highest level - a formula is used, a Palgrave index, that has no basis in index number standards and practice and is recognized as biased.

Rents and Domestic services indices

- We raise concerns regarding the quality of the data used as inputs to both the Rents and Domestic Services indices – basic headings that together account for one-quarter of the weight of the index.

- Inappropriate weighting and use of poor quality rent data have resulted in implausible and statistically inaccurate indices. These calculations should be corrected immediately. We estimate that correcting both the formula and index errors would have the combined impact of increasing the rents index for Geneva by up to 20% and increasing the PAI by approximately 4%.

- Inappropriate weighting and inconsistent editing is used to calculate the Domestic Services Index resulted in implausible and statistically inferior results. These calculations should be rectified immediately. We estimate the combined impact of these improvements would increase the Geneva PAI by close to 0.5%.

Medical insurance index

- We raise concerns regarding the compilation of the medical insurance index. No attempt has been made to standardise the respective baskets of the insurance policies included in the calculations. This is very problematic as it undermines any argument that like-for-like services are being compared.

Education index

- Errors identified in calculation of the education index, arising from simple clerical mistakes were identified and acknowledged by ICSC. They estimated the impact of correcting this error would increase the Geneva PAI by less than 0.1%. While this does not have a significant overall impact, it should be corrected nonetheless. Furthermore, this error raises concerns regarding how errors are dealt with in general and how revisions are incorporated and disseminated.

- When the impact of changes to the rent index, the domestic services index and the education index are taken into account, we estimate the reduction to the published May 2017 PAI will fall to less than -4%. This is well below the traditional 5% threshold for change.

More generally

- When the impact of changes to the rent index, the domestic services index and the education index are taken into account, we estimate the reduction to the published May 2017 PAI will fall to less than -4%. This is well below the traditional 5% threshold for change.

- Although not discussed in any detail in this report, we have concerns with the compilation with a number of other basic-headings, including rail fares and air fares.

- The focus of this review has been a comparison between Geneva and New York in 2016. As many of the issues raised are systemic in nature, we believe that our findings have direct relevance for other duty stations.

ANNEX I

Section 2 - Issues to be addressed immediately

2.1 Calculation of rent sub-index

19. By far the most important basic heading for any immediate review of cost-of-living calculations is the rental component given its weight of almost 23%. Furthermore "rents" are one of the most challenging and difficult elements to measure in a cost-of-living comparison. The challenges arise from a general lack of 'like-for-like' properties in similar locations/neighbourhoods being routinely available to price on a consistent basis. Other complexities such as different rental markets in different locations, local rent variations due to proximity to particular amenities or sites, different traditional property sizes and characteristics, ambiguities in how size is defined in each market and different types of legal contracts in each location, combine to make pricing comparable rental properties across time and space extremely difficult.

20. Leading statistical offices in the area of rental price measurement implement advanced estimation methods based on comprehensive datasets. As it happens, both Switzerland and the United States are considered among the front-runners internationally in this regard and both are considered to have high quality rent indices, properly tracking the evolution of changes in rental prices over time. However, such sources do not always allow for direct price comparisons across countries, either because average rental prices are not produced (only estimates of change over time) or there is incomparability between countries in the methods used.

21. In order to fill the gap in comparable average rents data, the ICSC post adjustment system, and the Co-ordinated Organizations (NATO, ESA, EUMETSAT, ECMWF, Council of Europe, OECD) & Associated Organizations have relied upon an annual survey of estate agents in different duty stations of the EU and the UN HQ duty stations. This survey is completed by the International Service for Remuneration and Pensions (ISRP) which provides rent data to ICSC and Co-ordinated Organizations for their respective purposes in adjusting remuneration and pensions across locations. The methodology underlying the ISRP survey is outlined in documents published by Eurostat and the ISRP[13].

22. The ISRP provides the average rental data used by both the ICSC and Eurostat (among others) for their calculations for various post adjustment type purposes. The compilation methodology uses a Fisher price index formula; a Fisher index is a geometric average of a Laspeyres price index (using New York weights) and a Paasche price index (using Geneva weights). The weights are based on information provided by staff through the staff expenditure surveys.

23. There are several difficulties with the estimated rent index as calculated for Geneva. These are outlined below.

[13] See, for example, the results of the 2016 exercise at: http://ec.europa.eu/eurostat/documents/6939681/7243182/REV_Booklet_2017_rents_2016_e_FINAL.pdf/87f4aa0c-a25a-400a-8a6e-e6caca2cbe99

2.1.1. Aggregation formula error

24. Aggregation takes place in two stages. Initially price relatives are estimated for Geneva (GNV) to New York (NY) for six categories of dwellings (studios, one-, two-, and three-bedroom apartments, semi-detached, and detached houses). The formula used at the higher level is a Fisher index and is used to derive a weighted rental value (price) index across the six categories of dwellings to form the basic heading index for "Rental for housing." The Fisher index is calculated as the geometric average of Laspeyres and Paasche, as outlined in Appendix 1.

25. However, in the ICSC case, Laspeyres and Paasche are ill-measured and substantially biased since quantity weights instead of expenditure weights are attached to price relatives. Although described as a Fisher index, the ICSC used an incorrect formulation of this index for the rent index: it is wrongly calculated. The aggregation formulae used in calculations of this type are well defined and established, namely: Laspeyres is a weighted arithmetic mean of Geneva to New York price comparisons that uses weights from the 'base' – in this case New York – duty station. Paasche is a weighted harmonic mean of Geneva to New York price comparisons that uses weights from the 'base' – in this case Geneva – duty station. Fisher is the geometric mean of the Laspeyres and Paasche indices. The weights in this case reflect the relative importance of different types of property within each index based on the property types rented by UN staff: studio apartments; 1 bedroom flats; 2 bedroom flats; 3 bedroom flats; non-detached houses; and detached houses.

26. An essential element of these calculations is the use of appropriate weights. A Fisher index requires, by definition, the use of relative expenditure weights (amount of spending on each type of property). **The ICSC use relative quantity weights** (number of staff living in each type of property) in their calculations. The resulting index is ill-defined with little meaning. We stress that the use of relative quantity weights is not due to an absence of expenditure data. It appears to be simply an inappropriate application of the formula.

27. For the 2016 round of surveys the ICSC quantity weights (as confirmed by ICSC and in ICSC documentation) were incorrectly input into the Laspeyres and the Paasche index formulae which are designed to use expenditure weights. It should be noted that this point is explicitly addressed in the Eurostat/OECD PPP manuals (2012, p.146), where they note that expenditure weights should be used with these formulae.

28. The scale and direction of the statistical bias associated with this incorrect approach can be approximated by estimating expenditure weights and using these in the calculation. Different approaches can be adopted to estimate expenditure weights[14] but the most technically consistent and transparent

[14] The authors' attempted alternative calculations of expenditure weights using just the data presented in Annex V of the reports (ACPAQ/39/R.3 and R.4) of the results for Geneva and New York for 2016, or a mix of the data from Annex V and Annex III, which could theoretically be consistent with the methodology approved for the rents calculation. For any of these alternatives, the impact on the overall Geneva PAI remains around 1%. The main figures quoted in the body of this report represent the

approach yields the results shown in Table 2.1 below. Readers will note that the impact is not insignificant - owing to the importance of the rental weight, this alternative estimation on its own yields approximately a 1.3% increase in the Geneva PAI.

Table 2.1: Rent index for Geneva - ICSC estimate and corrected estimate
(Replacing quantity weights with estimated expenditure weights)

Index Types	ICSC Index	Alternative Index using expenditure weights
Laspeyres	95.85	98.82
Paasche	89.22	97.64
Fisher	92.47	98.23
Absolute difference in rents index		+5.76
Percentage difference in rents index		+6.22

Source: ICSC (ACPAQ/39/R.4) and authors own calculations

29. The origins of this error appear clear. Up to 2010 the ICSC used common dwelling class expenditure weights for different property types. These common weights had been in place since early 1995 approximately (source Eurostat/IOS)[15] and used in calculation of the rent index for all Group I duty stations across all survey rounds since that time.

30. Following the 2010 round, a decision was made to change the methodology for calculating the rent index. Specifically, it was decided to move from common weights to duty station specific weights, which would be calculated using data collected from the staff expenditure surveys in order to make the calculations more representative of actual accommodation patterns in each duty station.

31. However, while the source of and the type of weights changed, the calculation formula was not appropriately updated in parallel, i.e. the calculation continued to use Laspeyres and Paasche index formula, the same calculation formula used by Eurostat, but that require expenditure weights (in other words the amount of expenditure spent on different property types).

32. Unfortunately, the ICSC used quantity weights (i.e. the number of staff living in different property types) resulting in a statistically invalid index. We assume this error arose as the ICSC did not fully understand the implications of changing from expenditure to quantity weights but this can only be confirmed by ICSC. The difference in weights applying to different property types is shown in Table 2.2 below.

33. The impact of incorrectly using quantity rather than expenditure weights in a formula that

calculations using only the data from Annex V which is the technically most consistent and transparent approach.

[15] As described in ACPAQ/33/R.2

requires expenditure weights is simply illustrated for the Paasche index: it gives higher weighting to more expensive properties (in the case of Geneva - houses). Using quantity weights as if they were expenditure weights exaggerated the impact of change in the rent parity between Geneva and New York causing the ICSC to incorrectly identify this as the impact of the change in methodology. A more detailed analysis is in Appendix 1. *A priori*, the direction and extent of statistical bias between quantity- and expenditure-share weighted price relatives are difficult to gauge. Thus, the use of quantity-weights is not only wrong, but also the nature and extent of the error they induce is difficult to predict and thus can be volatile and unpredictable if used for further rounds. The use of quantity-share weights is both biased and unreliable in measurement formula.

Table 2.2: Difference in weights for Geneva rent calculation

	2016			2010
Property Type Weights	Quantity weights	Expenditure Weights	Absolute Difference	Common Weights
3 bed flat	21.0	22.5	1.5	17.1
2 bed flat	27.8	23.1	-4.7	29.4
1 bed flat	18.7	11.3	-7.4	8.8
Studio	4.5	1.8	-2.7	1.0
Non-detached house	14.0	18.3	4.3	27.9
Detached house	14.1	23.1	9.0	15.8

Source: ICSC (ACPAQ/39/R.4) and authors own calculations

34. Hence the calculation of the rent index as currently compiled by the ICSC is not in line with the 'approved' methodology as the calculation has used an inappropriate combination of index formula and weights. The results shown here are approximations and full estimates of the impact would need to be estimated through ICSC's calculation system. Nonetheless, the authors believe that any recalculation will show a statistically significant negative statistical bias in the rent index for Geneva as currently compiled by the ICSC (see Appendix 1 for further discussion). In addition, as also discussed in Appendix 1, the authors believe the choice of the Fisher index, while not invalid, could usefully be discussed for future rounds as preferred alternatives could be identified.

2.1.2. Average rent trends:

35. The average rents are estimated by the ISRP annually through a survey of rental agencies in all respective duty stations. This involves a senior person in each agency being asked to provide an estimate of average rents for a range of different property types across the different neighbourhoods included. A heavy emphasis is placed by ISRP on trying to ensure that a like-for-like comparison is achieved so that results are comparable over time (e.g. similar quality of property, similar neighbourhoods, etc.). For the period 2010 to 2016, few changes to the neighbourhoods included were made and these took place

incrementally. Table 2.3 below shows the range of neighbourhoods included between 2011 and 2016 for both Geneva and New York. For each new location, the year of introduction is included in brackets.

Table 2.3: Neighbourhood locations used in 2016 calculations and Year of introduction

Neighbourhoods used in 2016 calculation	
Geneva	New York
Areas covered up to 2011	*Areas covered up to 2012*
Petit Saconnex	Manhattan (bounded by Yorkville, Park Avenue, Gramercy, East River)
Grand Saconnex	Westchester (Scarsdale, White Plains, Mamaroneck, New Rochelle, Larchmont)
Champel	Queens (Forest Hills)
Florissant	New Jersey (Closter, Fort Lee)
Malagnou	
Chenes-Bougeries	
Areas introduced since 2013	*Areas introduced since 2013*
Eaux-Vives (2013)	East Manhattan (inlc. Roosevelt Island, Waterside Complex) – flats
Chene-Bourg (2013)	Forest Hills (Queens) – houses and flats
Centre-ville/Plainpalais (2015)	Brooklyn (Brooklyn Heights, Carroll Gardens, Cobble Hill, Boerum Hill, Fort Greene, Park Slope) – flats
Nyon (2015)	Westchester – houses and flats
Mies (2016)	

Source: ISRP/Eurostat Annual Reports on Current Market Rents

36. In the case of Geneva, the evolution is clear – additional areas were added in 2013, 2015 and 2016. By and large, these new introductions would not be expected to have any substantial impact on the measurement (e.g. Nyon area would be cheaper that Geneva). This is true because the average level of rents and rental trends in the newly introduced neighbourhoods are unlikely to be very substantially different than those of existing neighbourhoods. In addition, other areas stayed in the basket meaning all introductions were only incremental, thus minimising the impact on overall averages (e.g. introducing one new neighbourhood in 2016 would have low impact even if that neighbourhood had different averages to the other 10 areas already covered).

37. To take one example, Eaux-Vives has higher rents per square metre than other neighbourhoods in the basket according to official rent statistics from the Cantonal Statistical Office of Geneva. Accordingly, averages measured in 2013 could, in theory, have been higher than if the same basket as 2012 had been used. On the other hand, Plainpalais is reported as having similar rents to the average in Geneva, thus not creating any expected change when it was introduced in 2015.

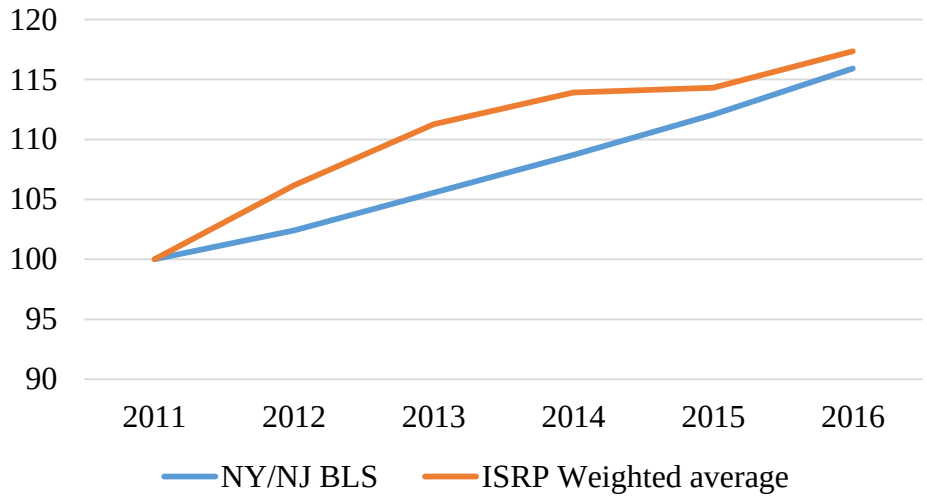
38. In New York it is less clear. The specified areas changed in 2013 but have been stable since, however there are some overlaps so the nature of the change is slightly less clear – however, indication from ICSC is that areas with cheaper properties have been introduced in recent years.

39. The message behind this is that, even allowing for the changes in neighbourhoods, the change in average rents measured by ISRP over time should be expected to track underlying trends in rents reasonably well over time. To assess this, we can evaluate the trend in ISRP data against externally available data sources for New York and Geneva.

40. Figure 2.1 presents a comparison of how rents for different property types in New York have evolved between 2011 and 2016 as reported by the ISRP and the Bureau of Labour Statistics rent index.

To simplify presentation, the six different property types used in the ISRP are aggregated into a single weighted average[16]. The overall increase in this weighted average was 17.3% which is slightly higher than the 15.9% increase recorded by BLS. Figure 2.1 shows that the two series were relatively consistent. Indeed, if anything, it seems the ISRP data may have overstated rental increases up to 2014, but following the change in methodology to introduce areas with lower rents (as understood by ICSC), the gap narrowed slightly.

Figure 2.1: New York rents trend
BLS (CPI) rent index compared with ISRP rent index
2011 - 2016 (2011=100)

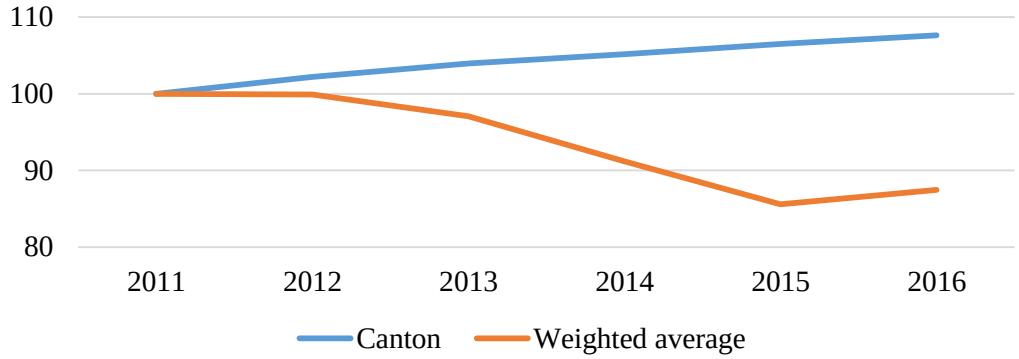


Source: BLS (NY/NJ Rental Index) and authors own calculations based on data in ACPAQ/39/R.3

41. The situation in Geneva is very different however. The Geneva rents series compiled by the ISRP is very different from the rent series published by the Cantonal Statistics Office for Geneva. Figure 2.2 shows how these two series have evolved over the same period, 2011 - 2016.

[16] If the individual price index for each property type is presented, the message is not changed. The individual ISRP series are distributed within a range above and below the BLS trend with the lowest increase being 9% over the period and the highest increasing by 27%.

Figure 2.2: Geneva rents trend
Canton of Geneva rents index (all rentals) compared with ISRP rents data (weighted average)
2011 to 2016 (2011=100)



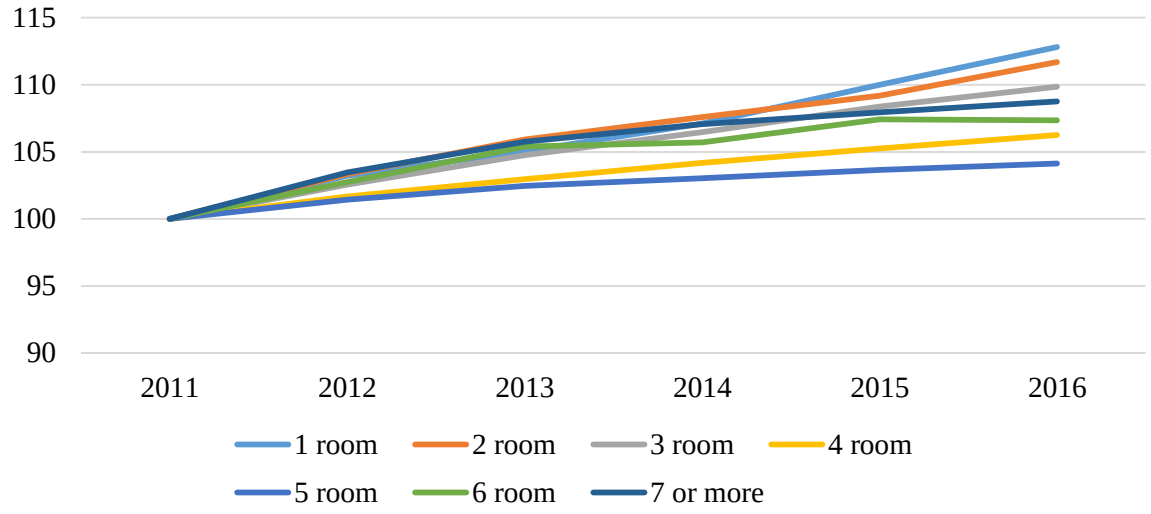
Source: Statistical Office of Canton de Genève and authors own calculations based on data in ACPAQ/39/R.4

42. The official series from the Canton of Geneva[17], which it must be recalled is recognized to be of high quality and very robust, shows a relatively steady increase over the period with a total increase in rent prices of 7.7% between 2011 and 2016. By contrast the weighted average ISRP index declines by a significant 12.6%.

43. The deviation between the ISRP and Canton of Geneva price series is even more striking when the disaggregated series are examined. Looking first at the official series published by the Cantonal Statistics Office for Geneva we can see a relatively narrow range of movement across properties of different size (see Figure 2.3). All property sizes increased over the period, by a minimum of 4% and a maximum of 13%. The official statistical office also produces very detailed disaggregated data by neighbourhood, metre squared and various other characteristics. All indices show trends within a relatively narrow range and again all indices reviewed increased between 2011 and 2016 by on average 8%.

[17] Data from the Swiss National Bank has a slightly different trend but similar increase overall over the period

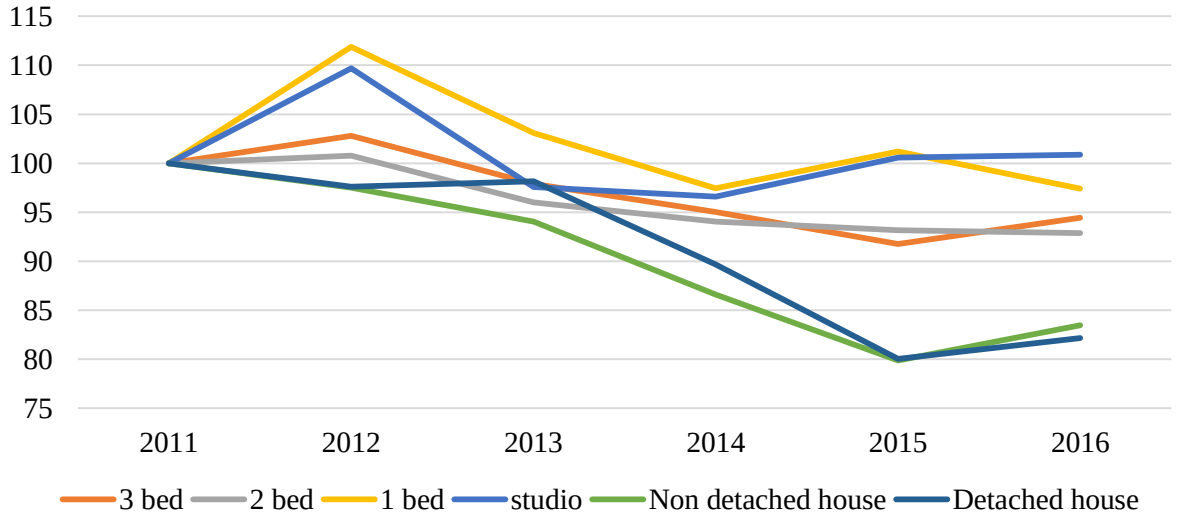
Figure 2.3: Official Geneva rent index by number of rooms
2011 - 2016 (2011=100)



Source: Statistical Office of Canton de Genève

44. A review of the ISRP data shows a very different picture. Compared with the official statistics, Figure 2.4 shows, not only a very different direction of trend (downward) but also a much more volatile and inconsistent evolution of prices in the ISRP data. Worryingly, the ISRP trend bears no resemblance to the official series (unlike the New York situation) and consequently questions must be asked regarding the quality and plausibility of the ISRP index as compiled for Geneva.

Figure 2.4: ISRP rent index by property type
2011 - 2016 (2011=100)



Source: ICSC (ACPAQ/39/R.4)

45. Initially some of the series increased, quite significantly in the case of smaller properties. However, from 2012 on the series started to dip substantially. By the end of the period only studios showed a level of rents similar to the starting point in 2011. For the other property types a negative trend is identified and in the case of non-detached and detached houses, the downward trend between 2013 and 2015 is dramatic (close to a 20% drop in 2 years for each series). This volatility and very dramatic falls in rent are implausible and inconsistent with how property markets generally behave. It is most unlikely that the change in neighbourhoods could account for these types of changes. For example, no changes in the neighbourhood basket was introduced between 2013 and 2014 and yet between those two years the ISRP rent index fell by approximately 6% on average while the official series increased by 1.2%.

46. The volatility of the ISRP data and the lack of any correspondence to the more robust official series raises serious questions regarding the reliability and credibility of the ISRP survey data for Geneva. As noted, this is in marked contrast to the situation illustrated for New York where the ISRP series was relatively consistent with official series. Furthermore, another additional external benchmark is available: the US State Department increased their annual Living Quarters Allowance granted to diplomatic staff based in Geneva between 2011 and 2016[18].

47. Through discussions with the ICSC, the authors attempted to ascertain if there is any valid reason why the ISRP data should diverge from official series. The ICSC responded that is their understanding that trends in the ISRP data should reflect genuine trends in rentals and this has been the basis for ICSC statements that rents have fallen in Geneva over the period. Worryingly however, there appears to have been little post-hoc external coherence testing to ensure that the results are plausible and explainable. The Eurostat-OECD Manual (2012, p.150) stresses the importance of conducting validations for consistency over time and plausibility. Review of ISRP documentation confirms that achieving like-for-like comparison, and thus reliable trends in rentals, is a key objective of the survey process even if some change can occur over time. In the case of Geneva, very serious doubts exist as to whether this has been achieved.

48. The main suggested reason for the discrepancy between the data sources evidently lies in the differences in their methodology. One relies on subjective assessments or estimates provided by a relatively low number of rental agencies (ISRP) whereas the other is based on a comprehensive dataset of real transaction rents. While the subjective approach can yield usable estimates as demonstrated in New York, in the Geneva case it clearly has not done so and must be considered of inadequate quality for such as sensitive purpose as salary adjustment.

49. Given the evident unsuitability of the ISRP data, this begs the question as to what the index for rents might have looked like if more appropriate data had been used. One possible alternative is to apply the trends from the official series to the base ISRP data from 2011 (this is the base in the sense as a six-

[18] The change in the Geneva rental index is well below the change in the Annual Living Quarters Allowance granted to diplomatic staff in the US Permanent Mission to the United Nations Office and other international organizations in Geneva. According to the data disseminated by the US Department of State (https://aoprals.state.gov/Web920/location_action.asp?MenuHide=1&CountryCode=1109), the Annual Living Quarters Allowance was increased by 20-36 per cent between 2011 and 2016, amounting in 2016 to US$ 76,800-93,300 per year.

year moving average is calculated covering 2011 to 2016). This at least presents a more plausible time series of the average rents data.

50. The series presented in Figure 2.3 were applied to the different property types from the ISRP (e.g. the series for '1 room' was applied to update the ISRP estimate for a studio, 2 room for 1 bed flat, etc.). This yielded the estimated average rents presented in table 4 below.

Table 2.4: Updated average rents in Geneva by property type (CHF)

Year	3 bed flat	2 bed flat	1 bed flat	Studio	Non-detached house	Detached house
2011	3,950	3,084	2,113	1,406	5,225	6,520
2012	4,016	3,163	2,184	1,451	5,300	6,723
2013	4,067	3,231	2,238	1,477	5,354	6,884
2014	4,115	3,283	2,274	1,506	5,383	6,938
2015	4,158	3,342	2,307	1,547	5,416	7,022
2016	4,197	3,388	2,360	1,586	5,441	7,047

Source: Authors own calculations based on official statistics compiled by Statistical Office of Canton de Genève

51. This creates a substantially different six-year moving average as shown in Table 2.5 below. The six-year moving average rent for Geneva increases by between 6.6% and 19.5% depending on property type. The greatest differences are observed for houses. This dispels any possible argument that any issues with year-to-year measurement basis are balanced out when a six-year average is used.

Table 2.5: Comparison of Alternative and ISRP/ICSC six-year moving average rents for Geneva (CHF)

6 Year moving average	3 bed flat	2 bed flat	1 bed flat	Studio	Non-detached house	Detached house
Alternative	4,107	3,277	2,268	1,513	5,372	6,901
ISRP/ICSC original	3,777	2,925	2,127	1,409	4,590	5,777
Percentage difference	*8.7*	*12.0*	*6.6*	*7.4*	*17.0*	*19.5*

Source: ICSC (ACPAQ/39/R.4) and authors own calculations

52. Integrating these new average rent estimates into the overall calculations of the comparison between New York and Geneva has a statistically significant effect as illustrated in Table 2.6.

Table 2.6: Alternative rent indices – Geneva to New York

Index Type	Index number	Diff vs ICSC estimate (absolute)	Diff vs ICSC estimate (%)
Index (Original ICSC)	92.47	N/A	N/A
Index with corrected aggregation formula	98.23	5.76	6.22
Index with alternative average rents but incorrect aggregation formula	103.41	10.94	11.83
Index with alternative rents and corrected aggregation formula	110.68	18.21	19.69

Source: ICSC (ACPAQ/39/R.4) and authors own calculations

53. Simply put, correcting both the formula error and the implausible input rents data increases the rental index from 92.47 to 110.68. It is the view of the authors that this represents a more plausible estimate and remains faithful to the intention and methodology of the rent index.

54. It could, of course, be argued that the rents index for New York could similarly be updated, but as demonstrated in Figure 2.1, this would only have a minor impact, as ISRP and the BLS New York Rental Index give quite consistent average rents for New York. In fact, as the ISRP averages increased by more over the period than the BLS index, it would only have the impact of increasing the Geneva relativity slightly further.

55. A critical point to note is that while use of ISRP data is 'approved', this approval or any other cannot remove the requirement and responsibility to ensure the data used are *'fit for purpose'*. Margins of error exist which can be tolerated to some extent, in so far as they do not evidently statistically bias the estimate of interest, as is the case for New York. For Geneva, however, this basic condition has not been met. The volatility and extreme declines identified by the ISRP cannot be explained or reconciled with the publicly available official statistics. In the opinion of the authors, these data do not meet the minimum quality standards required and cannot therefore be considered robust or reliable or free from significant statistical bias.

56. The ICSC has recognised the need to adapt to situations where results of the 'approved' methodology did not yield plausible results. For example, the ICSC and ACPAQ have agreed that supplementary price collections were needed to improve comparisons between duty stations where ECP price collection had taken place. This decision was taken as the ECP data (the use of which had been approved by ACPAQ) did not yield plausible comparisons with the ICSC data from New York. This decision was appropriate. Such flexibility may be required from time to time when the approved methodology generates unusable results. This is clearly the case for the rents data as demonstrated

above[19]. How to avoid this situation arising again in future can be a subject of further discussion, but the existing index must be corrected before the current ICSC results can be considered *'fit for purpose'*.

57. Using the results from Table 2.6, the authors have estimated the impact on the overall Geneva PAI. Replacing the original ICSC index (92.47) with the superior alternative (110.68), the overall PAI for Geneva would be increased by approximately 4.1% as shown in Table 2.7. It must be noted these estimates are indicative – proper estimation of the updated series would need to be computed by ICSC using October 2016 as the base and updated to May 2017 but we would not expect a significantly different result.

Table 2.7: Impact on Geneva PAI using alternative rents calculations

May 2017 Geneva PAI (as published by ICSC)	167.7
May 2017 Geneva PAI (with rent correction)	174.6
Estimated underestimation due to rents (%)	*4.1*
May 2017 Pay Index (as published by ICSC)	181.1
Difference in Pay Index (%) (prevailing/corrected)	*3.72*
Difference in Pay Index (%) (original - as published by ICSC)	*7.99*

Source: Authors estimations and ICSC May 2017 Post Adjustment Circular

58. Finally, it could in fact be reasonably argued that the difference should be even greater. For example, in their guidance for interviewers, the ISRP instructs them to collect 'the range of rentals within which the most recent contracts have fallen' (2017, pg.9). This implies that a more appropriate comparison would be with a series of new rentals rather than all rentals which includes existing contracts. The official series for Geneva show that the inflation in new rental contracts is far higher than for existing contracts and thus could create a significantly different outcome. Furthermore, had these higher average rents data been used to calculate new expenditure weights it would have increased the index even further again as additional weight would have been given to houses. However, such issues can be considered in a more in-depth review of the current calculation and methodology. For the current purposes, the authors are of the opinion that, the data in Table 2.7 is the minimum level of adjustment required to the Geneva PAI due to the problems identified with the rents calculations.

[19] It should be noted that significant concerns regarding the rents data were raised by representatives of staff associations and management of Geneva-based organisations at the 39th session of ACPAQ. However, the analysis presented above was not available at that time.

2.1.3. Summary of problems identified with rents calculation:

- A formula/weighting error was identified resulting in an underestimation of the rent index. We estimate this under-estimate at 6.2% which in turn accounts for approximately 1.3% of the overall PAI for Geneva.
- The ISRP rents data used in the Geneva calculation for the period between 2011 and 2016 are problematic. They are volatile, implausible and inconsistent with other publically available official rent statistics. Replacing the series used with official series yields a rent index increasing by 12% versus the ICSC calculated index.
- Neither of these issues impacts on estimates for New York as (1) the ISRP rents data for New York is consistent with official BLS rent data and (2) the formula/weighting error only occurs when the comparison between New York and Geneva is calculated.
- The combined impact of correcting the formula/weighting error and using a superior rent index increases the Geneva rent index by 19.7%. This in turn increases the PAI by 4.1%.

2.1.4. Other issues of note with rental data:

59. There are a range of additional issues which the authors consider to be worthy of further review with regard to the rent calculations, but which do not have any clear immediate impact or where no alternative calculation could be made in the time available. These include:

- Size adjustment factors are used but only for 3 of the 6 property types. These effectively adjust the rents in New York to account for typical larger property sizes. However, no such adjustment is made for smaller properties (studio apartment, 1 bedroom flat or 2 bedroom flat). The source of the size adjustment factors are size bands used by ISRP for their price collection. However, these bands are arbitrary, very wide and have no obvious basis empirically. These size adjustment factors have a significant impact on the results for rents. Insufficient information is available to suggest an alternative set of thresholds, but it is suggested that a review take place to establish if they continue to be valid or whether more appropriate alternatives could be developed.
- The indices for updating used by the ICSC are the all-items CPIs for US (cities) and Geneva respectively. In both cases more appropriate indices are available (rents indices for New York/New Jersey/Long Island and Geneva respectively). These would be more meaningful in application. It can, however, be noted that for the 2016 round, this would not have had any tangible impact on the index as the difference between the CPI in Geneva and US over the period was similar to the difference in the rent indices for the two locations. In that sense, it can be considered as a subject for future consideration.
- Given the inappropriate results generated by the ISRP, a thorough review of possible sources of rental data is required. This has been discussed at previous meetings but never before has the outcome shown such statistical bias and flawed results. This creates a new imperative to find and use alternative data sources, at least in the Geneva case and quite possibly for other duty stations. Many possible approaches can be imagined such as modelled estimates based on large available databases. However, the starting point of such a change is a discussion on the intended output from such an exercise to avoid a

repeat of the situation where a methodology that is seemingly appropriate generates implausible and inaccurate results.

- It should be noted that the 2016 round of surveys is not the only use of the ISRP data. These data are used on an annual basis to update the Geneva and New York PAIs. As noted above the impact of this has been minimal on New York. However, as demonstrated, the impact in Geneva has been significant and has had a significant downward effect on the Geneva PAI over a number of years. This in part explains the gap between the Geneva Pay Index and the Geneva PAI which has been used as a justification for the inference that Geneva has been 'overpaid' vs actual relative cost of living[20]. However, it should be made clear that while this did not in fact influence actual pay levels during the 2010 to 2016 period, it could have substantial impact in the future and should be borne in mind in any consideration of updating the source of rent data.

2.2 Calculation of Domestic Services sub-index

60. Domestic services accounted for 2.5% of the basket in the cost-of-living comparison for 2016 and so is an important consumption expenditure item. As with the rent index, the pattern of the domestic services index raised questions regarding the plausibility and quality of the calculation. Between 2010 and 2016 the domestic services index swung from 118 (i.e. prices were 18% higher in Geneva than New York) to 83 (i.e. Geneva prices were 17% lower than New York). This significant swing accounted for almost 0.9% to the total pay cut (versus a situation where the index had stayed the same). The 2016 index of 83 was compiled from 7 sub-indices weighted together in two stages as outlined in Table 2.8 below.

[20] ICSC/84/R.8 paragraph 103

Table 2.8: Domestic services index calculation (ICSC)

	Cost (US$)		Geneva Weights	New York / Geneva relative
	New York	Geneva		
Full-Time (monthly wage)				
Maid	1,422.5	2,181.9	16	1.534
Babysitter/childminder	3,044.2	2,693.2	84	0.885
Cook	-	4,072.0	3	-
Other	1,625.0	1,851.8	20	1.140
Sub-index				1.014
Part-Time (hourly wage)				
Maid	35.7	28.0	1,302	0.784
Babysitter/childminder	24.9	23.7	373	0.954
Cook	41.8	27.4	42	0.655
Other	29.2	24.2	102	0.828
Sub-index				0.818
Total Index				0.830

Source: ICSC (ACPAQ/39/R.4)

61. The source of the domestic services data are the staff expenditure surveys. Staff were asked to complete a table detailing the different domestic services they pay for. A snapshot of the relevant section of the questionnaire is shown below in Table 2.9.

Table 2.9: Excerpt from ICSC Family Expenditure Survey, 2016

			Cash wage			**Other monthly related costs**	
	Days per month	Hours per day	Frequency	Amount	Currency	Amount	Currency
Housekeeping services							
Cooking services							
Babysitting/Child-minding services							

Source: ICSC Survey on Household Expenditures 2016

62. Some additional information was also captured on other domestic services but the majority of the data used in calculation of the index was generated from the section shown in Table 2.9. The notes provided to aid staff in completing this table were minimal which resulted in a significant number of staff making various errors concerning frequency, hours and amount. An inspection of the data shows that some staff did not complete all the necessary fields (e.g. hours per day) but for most staff sufficient information appears to have been provided to allow estimations to be derived, albeit with some subjective judgement required during data processing to correct obvious errors.

63. Discussion with the ICSC indicated that domestic services is not an area where any substantial methodological change took place between 2010 and 2016 that would have been expected to influence calculations. The questionnaire was changed to consolidate the information collected into one section but this does not appear to have altered the nature of the data collected.

2.2.1 Data processing and editing:

64. During the mission to New York a relatively substantial issue was identified with respect to the editing procedures applied to the Geneva and New York dataset. The general approach applied across different item headings was to identify outliers or errors using a set of standard editing rules supplemented with judgement decisions made by the statistician editing the data. As outlined in ICSC documentation high and low outlier thresholds were identified programmatically through relatively standard statistical methods (based on a number of standard deviations from the mean). In this way observations were identified as valid or not valid. Editing could also take place for other reasons, such as removal of prices as the item did not match specification etc.

65. In the case of domestic services an inconsistency was noted with regard to the treatment of outliers between Geneva and New York. Table 2.10 below shows the approximate minimum, maximum and mean of price observations used in the average price calculations for hourly wages for part-time maids as an illustration of this.

Table 2.10: Estimates of hourly wages of part-time maids used in the calculations,

New York and Geneva, 2016

	Min (approx.)	Max (approx.)	Median	Mean (ICSC)	Mean/Median ratio
Geneva	USD11	USD60	USD25.30	USD27.96	1.105
New York	USD5	USD240	USD26.68	USD35.67	1.337

Source: Authors review of ICSC database (provisional and subject to update)

66. While both databases contained both very high and low values (as reported by staff) there appears to have been an inconsistency in the editing procedures applied to the files for Geneva and New York. This is most easily illustrated by examining the maximum values included in the calculations for both duty stations. Up to a maximum of approximately US$60 per man-hour was allowed in the case of Geneva whereas a maximum of US$240 per man-hour was allowed in New York. The impact of the two different upper thresholds applied in the two duty stations is immediately obvious: a min/max ratio of 1: 48 in New York compared with a min/max ratio of 1: 6 in Geneva. A similar pattern is also evident when comparing the mean and median averages in both duty stations. In Geneva, the mean is approximately 10% higher than the median, whereas in New York the mean is 35% higher than the median. Despite testing various common outlier detection methodologies (e.g. 2 standard deviations from the mean, IQR/Boxplot) the authors were not able to identify the specific approaches adopted and neither seemed to match the ICSC specified approach for other items in the basket. However, what can be said is that one approach appears to have been programmatic (in Geneva all values above CHF60 per hour were excluded) while there is no clear pattern in New York with 108 values in excess of $60 left in the calculations representing about 12% of all the values used to get the average price.

67. Without access to all the respondent notes and metadata the statistical justification for adopting an inconsistent approach is not obvious. But in our opinion, it is highly improbable there is anything there that would justify such a difference in treatment, and consequently we believe relative price difference between New York and Geneva is exaggerated and that the average price for New York is overestimated. Possibly different staff in ICSC processed the two different datasets, which led to quite different thresholds being applied. But if this is the case, common guidelines should have been available to ensure that editing statisticians applied the same approaches to both datasets.

68. A similar pattern was noted across the part-time categories in the other domestic services (babysitting, cook, other) whereby an upper threshold of CHF60 per hour appears to have been applied for Geneva while some higher values were allowed in New York.

69. Alternatives could be generated in various ways. A very simple alternative for illustration purposes is to apply a similar upper threshold to New York as Geneva, say $60 per hour. Table 2.11 shows the difference in results generated when such an approach is taken. It can be seen that all the indices increase, mostly substantially. The overall impact on the domestic services index is to increase it from 83.0 to 99.2 which in turn would increase the overall PAI for Geneva over 0.4%.

Table 2.11 – Impact of updated outlier threshold for New York

	New York Average Prices ($)		Price relatives	
	ICSC	**$60 Threshold**	**ICSC - Original**	**Using $60 threshold**
Part-time maid	35.67	28.36	0.784	0.986
Part-time babysitter	24.88	22.42	0.954	1.059
Part-time cook	41.80	33.09	0.655	0.827
Part-time other	29.24	28.17	0.828	0.859
Total part-time			0.818	0.990
Total domestic services			0.830	0.992
Impact on PAI				+0.4%

Source: ICSC ACPAQ/39/R.4 and authors own calculations

70. Another alternative is to use more scientific available outlier detection methods. The authors simulated results using an IQR/Boxplot method. Under that method all indices also increase with maids increasing by less but the indices for part-time cooks and 'other' increasing by even more than when a $60 threshold is applied. While the results differ slightly the same overall approximate impact on the PAI is shown (above +0.4%).

71. From our perspective, this type of statistical bias error is within scope, as it is an implementation issue and not covered by the 'approved' methodology. Correcting it simply ensures that the approved methodology is applied in a statistically appropriate manner. It is also of relevance to other duty stations as in this case one conclusion is that average prices used for New York are too high.

2.2.2 Aggregation:

72. The above only refers to the impact of statistically biased editing processes. In addition, an incorrect aggregation methodology is applied by the ICSC. The ICSC applies quantity weights to aggregate the lower level indices to the overall index for domestic services. In other words, 1302 staff members in Geneva used part-time maids and this was applied as the weight for the index for part-time maids. This approach is completely inappropriate to weight together

the indices between part-time and full time as they represent very different levels of expenditure. A person engaging a full-time maid will have far higher expenditure than a person engaging a part-time maid for a few hours a month. Weighting using quantities effectively gives those people equal weighting in the calculations.

73. Improved alternatives are known which, as with rents, involve using expenditure weighting. Taking the 'minimum change' approach, which is a weighted arithmetic mean (using expenditure weights estimated by the authors) would generate an alternative index of 86.66 using the ICSC original average prices. This is an increase of 3.7pp on the index of 83.0 calculated by ICSC. When the updated prices generated by the authors are also used a new result of 99.6 which would increase the Geneva PAI by 0.5% approximately. However, this formulation, as discussed is also not recommended as it creates a Palgrave index and valid alternatives are known (see Appendix 2 for more details).

74. While the above is a correction based on minimal change in the current approach, aggregation methodologies at the lower level for domestic services also require further consideration.

75. There is aggregation at the lower level that is the unweighted arithmetic (mean) of the prices provided for each different type of domestic service. These average prices are compared to form the price ratio for Geneva to New York average (a Dutot index). These price ratios are then weighted to give the overall index. By contrast for the in-area excluding housing items a ratio of geometric means is used (a Jevons index). The use of these formulas – Dutot versus Jevons - can yield very different results. The nature and direction of the differences is well established and accepted in the technical literature on index number methodology to be dependent on the difference in the variances of prices, in the two cities

76. Further, there are well-established axioms that a good index number formula should satisfy. This axiomatic or "test" approach finds a ratio of geometric means - the Jevons index - is preferable to a Dutot ratio of arithmetic means.

77. As such, we can, at first sight, conclude that ICSC is using a formula – a Dutot index - for its lower level averaging of domestic services that does not meet international standards. However, there is a complication. Unlike many other goods and services, property is sold/rented on an individual basis and for a Dutot index, when comparing matched properties, say over time, there is an implicit relative expenditure weight. Yet we are not dealing with matched properties. The review team found a complex issue. Yet this issue of matched properties and the complexity of the case for Dutot was not part of the dialogue. The use of Dutot as opposed to Jevons can make a substantial difference, yet there is little to no justification as to the varied use of the formula for lower level aggregation. The index for important components does not stand on well-reasoned, firm grounds in the calculation of its building blocks of average prices. A full review is required for what is a seemingly casual choice of varied formula.

2.2.3 Comparison with external data

78. Like the rent index, comparisons with publicly available official statistics and other commercial prices also throw serious doubt on the plausibility of the indices calculated by the

ICSC. The index for full-time maids is 153.4 (i.e. 53.4% higher in Geneva). The index for part time maids however is 78.4 (i.e. 21.6% cheaper in Geneva). While some differences could perhaps be explained it appears extremely unusual that costs for a full-time maid would be so much higher in Geneva, but for a part-time maid so much lower.

79. Table 2.12 below shows a comparison of data from the ICSC and BLS. It can be easily seen that the ICSC data does not resemble data from the BLS either in level or trend. Starting with 2011 the average hourly figure from ICSC for a part-time maid was 56% higher than the equivalent figure from the BLS Occupational Employment Statistics, while for babysitters the difference was 44%. However, by 2016 the ICSC part-time maid estimate had increased by nearly 55% and for babysitters there was an increase of 44%. By contrast both the BLS estimates showed a modest increase. As a consequence, by 2016 the ICSC estimates were more than double those from BLS.

Table 2.12 – Average hourly earnings for domestic services in New York*

	2011	2016	% Change 2011-2016
ICSC Part-time maid ($)	23.03	35.67	+54.9
BLS hourly maid ($)	14.81	15.69	+5.9
% Difference (ICSC/BLS)	+56	+127	
ICSC Part-time Babysitter ($)	17.27	24.88	+44.1
BLS hourly babysitter ($)	12.02	12.28	+2.2
% Difference (ICSC/BLS)	+44	+103	

*BLS New York estimates refer to the series for New York-Newark-Jersey City (series 35620)

Source: ICSC (ACPAQ/39/R.4) and BLS Occupational Employment Statistics

80. It is possible that the levels could be different at a point in time, for example because staff hired from an agency may not receive the full payment made by the client themselves. However, this does not appear to sufficiently explain the level of difference recorded either in 2011 or 2016. Even accepting that level differences could occur the difference in trend is extremely difficult to explain. There was no methodological change which would explain the extraordinary growth in hourly rates recorded for New York. The combination of these factors strongly suggests that the hourly rates for New York are exaggerated. Even the corrections proposed to the editing process would only partially deal with this as estimates from the ICSC would still remain close to double those from BLS in 2016.

81. Further to the above we can attempt to reconcile official data for Geneva with ICSC data. The data on earnings for Geneva is nowhere near as detailed as that available for the United States. Just taking the case of maids we can see a standardized monthly average level of earnings of CHF4,101 in 2014[21]. This is standardized based on the official methodology to an hourly rate of CHF25.02. We can note that this corresponds far more closely to the ICSC estimate for Geneva ($27.96 or approx. CHF27.62) than seen in the case of New York. Bearing in mind the timing difference between the estimates (2014 and 2016) and assuming any growth in the average between 2014 and 2016 one can conclude that the difference is relatively small.

82. Comparing the official Geneva data to the official New York data we get a comparison for 2014 of $27.35 for Geneva (converted to dollars based on the average 2014 exchange rate) compared with $15.69. This is a ratio of 1.743:1, in other words 74% higher in Geneva than New York. This compares reasonably closely with the ICSC estimate for the difference in costs for full time maids (53.4% higher in Geneva) at least in general order of magnitude. However, it is nowhere near the ICSC estimate for part-time maids (21.6% lower in Geneva).

83. In summary, the official data available suggests costs in Geneva for domestic services should be higher than in New York. The level and trend displayed in the ICSC data are difficult to reconcile with the implication that the levels recorded in New York for 2016 are potentially substantially overestimated.

Summary of key issues regarding domestic services:

- An inconsistent approach to editing New York vs Geneva data has resulted in a statistical bias, the impact of which appears to be a statistically significant underestimation of the index for Geneva. A more balanced editing approach should be applied to remove the identified statistical bias. This is a straight-forward operational issue and hence should be considered necessary to ensure appropriate implementation of ‘approved’ methodology.

- The aggregation approach of weighting relatives by quantity weights is not recommended and in breach of generally accepted practices. Those weights should be replaced by appropriate expenditure weights. We estimate that substituting appropriate expenditure weights into the calculation would increase the Geneva PAI.

- A full alternative calculation would require further work by the ICSC but the impact of the above issues is estimated as an increase for the Geneva PAI of between 0.4% and 0.5%.

- Notwithstanding these changes, doubts would remain regarding the quality of the index as external data would point towards an even higher index for Geneva and aggregation methods should also be subject to further review.

[21] Source: Statistical Office of the Canton de Genève

- Further to the above, which represents a 'minimum change' to existing approaches, aggregation approaches at both the lower (Dutot) and higher (Palgrave) levels should be changed as discussed in appendix 2.

- The issues highlighted are of relevance to other duty stations given the impact on average prices for New York and the overall calculation of the index. However, the extent of impact will differ across duty stations due to differences in weighting etc. Also, external data sources may often be lacking meaning elements of the analysis above may not be possible to complete.

An additional point of note:

84. It is not immediately clear why data from the staff expenditure surveys is used in this case – other than precedent. Given the design of the questionnaires, responses are prone to error which can significantly influence results. Furthermore, it is not clear whether the comparisons generated (even if the data were perfect) are the most appropriate source given the aim is to compare cost-of-living. On this point, the authors note that there is some inconsistency in the general approach adopted by the ICSC regarding their use of 'price comparisons' and 'cost comparisons' to derive relatives. This is an important distinction and one where the choice can heavily influence results. If a 'price comparison' approach was chosen for domestic services, a decision that would be more consistent with most other items included in the cost-of-living comparison, then prices could be perhaps be sourced from a selection of agencies or average hourly rates from official statistical series could be used, as highlighted above. Combined with appropriate expenditure weighting from the staff expenditure survey we would expect a more stable and potentially realistic series of results.

2.3. Other immediate issues

2.3.1. Calculation errors

85. During the mission to ICSC offices, the review team identified some basic calculation errors. For example, an error was identified in the calculation of the education index due to the incorrect exclusion of prices from the calculation for Geneva. The individual impact of this was relatively low (estimated to increase the PAI for Geneva than less than 0.1%). In a project as large and complex as the Cost of living surveys run by the ICSC, mistakes of this nature should not be surprising. Had we had more time, it is likely that we would have identified other mistakes. This is not intended as a criticism of the ICSC statisticians - one can expect data, statistical and process errors to arise in any large statistical project. While steps are taken to minimize such errors, they are never completely eradicated. But discussions following the detection of the error brought two issues into sharp focus:

a. In a complex statistical exercise like the ICSC which comprises of thousands of individual statistical estimates errors will almost certainly occur. These errors will vary in impact but some could have a significant overall impact. Thus, the final results of the ICSC, like most statistical indicators, are estimates at best and in no way analogous to audited accounts. The ICSC do not

quantify the relative precision of their estimates, but we can take it as understood that the published estimates are surrounded by a range of uncertainty (even when all identified errors are corrected). Consequently, the decision in 2015 to remove the 5 percent gap closure measure, which effectively acknowledged that the ICSC estimates had an unknown confidence interval, would appear to have been a particularly unwise one.

b. The ICSC statisticians noted there was no mechanism to correct the detected error. As we understand it, feeding the correction through the system would have had uncertain consequences but furthermore, as the revised salary scales had already been published, there was no way to undo the error in any real sense. In other words, salaries would continue to be adjusted, knowing that the statistics informing that decision were incorrect. Obviously, for practical reasons, there must be some cut-off point where a final decision is made, but equally, there must be a mechanism to correct calculations when significant errors are detected.

Section 3 - Issues to be addressed urgently

86. Section 3 outlines issues identified by the review team which we consider to be of very high importance and which we believe have statistically biased the 2016 results. Owing to lack of time or alternative information we have not been able to quantify the extent of the impact of these problems on the Geneva PAI.

3.1 Low numbers of price observations, weak averages and volatility:

87. Pricing data is collected for cost-of-living comparisons in a variety of ways. The large majority of prices are collected through place-to-place surveys such as those which took place in New York in June 2016 and Geneva in October 2016 which cover more than 300 specified goods and services. It should be noted that the number of goods and services included in the ICSC basket is far below the number included in the European Comparison Programme where 3,200 items are included[22].

88. Table 3.1 shows the number of prices initially collected by ICSC in New York (5,423) and Geneva (2,860) along with those used in the final calculations for the 2016 operation. Overall 38% of the prices collected in New York were dropped compared with 20% in Geneva[23]. The number of prices collected along with the percentage dropped varied heavily across major headings.

Table 3.1: Number of prices collected and used by major expenditure group, New York and Geneva, 2016

Expenditure category	NY Collected	NY Used	NY % dropped	GVA Collected	GVA Used	GVA % dropped
Food and non-alcoholic beverages	1,744	1,254	*28*	739	607	*18*
Alcoholic beverages and tobacco	329	213	*35*	165	149	*10*
Clothing and footwear	1,251	473	*62*	559	381	*32*
Housing, water, electricity, gas and fuels	73	54	*26*	55	39	*29*
Furniture, household equipment and maintenance for the house	715	373	*48*	314	229	*27*
Health	252	219	*13*	218	200	*8*
Transport	247	220	*11*	147	142	*3*
Communication	46	30	*35*	39	39	*0*
Recreation and Culture	376	268	*29*	303	260	*14*
Education	27	12	*56*	70	55	*21*
Restaurants and hotels	115	97	*16*	85	42	*51*
Miscellaneous goods and services	248	146	*41*	166	136	*18*
Total	5,423	3,359	*38*	2,860	2,279	*20*

Source: ICSC (ACPAQ/39/R.3 and R.4)

[22] In part, this relates to the need to cover non-consumer spending in the ECP programme but across all major groups, the ECP has far greater numbers of items than the place-to-place surveys run by the ICSC.
[23] Figures are taken from the documents provided to the 39th session of ACPAQ. Further rounds of editing since the 39th session mean the numbers of observations used in final calculations may be slightly different.

89. A higher number of prices were collected in New York than in other duty stations including Geneva. The ICSC explain this as necessary to ensure sufficient prices are available to allow like-for-like comparisons to be made with prices collected in other duty stations. This includes collecting prices for some items which may not be representative in New York but are representative elsewhere.

90. The data editing process worked in a number of stages and prices were excluded from calculations for a variety of different reasons. First stage editing included relatively standard outlier selection and exclusion (both high and low outliers) and exclusion of items because of un-matched specifications. The essential aim of this first stage was to remove prices considered invalid. A second stage of editing involved 'balancing' between duty stations. In this stage, prices and descriptions collected in New York and Geneva were compared. During this round, individual prices were excluded to ensure that only comparable items remained.

91. In the opinion of the authors, this process has led to low numbers of price observations being used in many cases which in turn has generated weak averages of unclear representativeness. In addition, the place-to-place surveys include a relatively low number of items for a process of this type[24]. Related to this, and exacerbated by issues with aggregation methodologies, we note an undesirable and inexplicable level of volatility between the 2010 and 2016 rounds across various basic headings. This can be argued to have statistically biased downwards the PAI for Geneva as explained further below.

3.1.1 Low numbers of price observations

92. As shown in Table 3.1 quite a high proportion of collected prices were not used in calculations, in particular from the New York database. For example, 62% of price observations were removed for *Clothing and footwear* in New York. In the case of Geneva 51% of price observations were removed for *Restaurants and hotels*. While this level of price exclusion could be considered surprising, it does not necessarily bias results as long as a sufficient number of representative prices are retained to enable adequate like-for-like comparison between locations.

93. However, it is observed that relatively low numbers of price observations have been used to generate average prices in both Geneva and New York across a range of items. For example, in Geneva over 40% of the items had 5 or less price observations used (and 9% have only one). By comparison for New York approximately 30% of items had 5 or less price observations used while 9% had only one price observation (see Table 3.2).

[24] For example, only one brand of beer, Heineken, was selected to represent the consumption of alcoholic drinks in restaurants and only one price observation was included for New York.

Table 3.2: Number of price observations used in 2016 calculations

Number of prices used	Geneva		New York	
	Number of items	% of Total	Number of items	% of Total
1	26	*8.8*	27	*9.1*
2	12	*4.1*	11	*3.7*
3	25	*8.4*	18	*6.1*
4	30	*10.1*	21	*7.1*
5	31	*10.5*	20	*6.8*
6	31	*10.5*	19	*6.4*
7	22	*7.4*	13	*4.4*
8 or more	119	*40.2*	167	*56.4*
Total	296	*100*	296	*100*

Source: Authors review of ICSC prices database

94. For some of the items, only one price observation has been included in the calculations, while very low numbers were included for other items. In some cases, this is expected due to a very specific item specification (e.g. a specific IKEA item). However, for other items it should have been possible to improve representativeness of the averages used. For example, only one price for the renewal of a monthly subscription to a fitness centre was collected for Geneva.

95. Reviewing the database indicated a relatively high number of prices collected which did not match with indicated specifications or where prices seemed unrealistically high or low with respect to the specification provided. This in theory is avoidable (or at least can be reduced) and for future rounds by improving item specifications.

96. The risk created by low numbers of price observations is the creation of weak averages which in turn can lead to unreliability of estimated averages and price indices. This unreliability will not necessarily be reflected in coefficients of variation calculated by ICSC as the risk is that the prices collected simply do not adequately represent spending patterns to form a reliable base for comparison. An additional risk is volatility of price indices between different rounds of place-to-place surveys.

3.1.2. An Illustration – long distance transport

97. To illustrate this point, let's examine the basic heading: long distance transport. In 2010 this index was 253.55 (in other words average prices in Geneva were 153% higher than in New York). In 2016 this fell to 81.77 (i.e. 18% lower in Geneva) – see Table 3.3. Such a change would imply dramatically different inflationary trends in Geneva and New York - such changes are not substantiated by official statistics in either jurisdiction. Put simply, this level of change

recorded is not plausible as a real world inflationary effect. Such volatility is surprising as these types of services are relatively easy to specify and there are a limited number of operators and pricing is typically very transparent.

Table 3.3: Price observations and price indices (2010 and 2016) for long distance transport

No. of prices used NY (15)	No. of prices used GVA	Price index 2010	Price Index 2016	% weight of basic heading
7	3	253.55	81.77	0.25%

Source: Authors review of ICSC database plus ICSC ACPAQ/39/R.4

98. In 2016, the ICSC specified a journey of 250km by train as the reference item. Initially just one journey matching this was included for Geneva giving just a single price observation. Similarly, in New York one journey was identified which approximately matched the specification and two prices were included (a value fare and a standard fare). Given the low number of observations, the ICSC added some additional shorter distance journeys (additional two prices in Geneva, additional five in New York). The ICSC adjusted prices for the different journeys to correct for differences in the length of journey covered.

99. Thus, three price observations (of which two did not match the initial specification) were used to estimate a comparable relative price between New York and Geneva. Given the volatility in the price index (253.55 to 81.77) one must question the reliability of the estimate. Furthermore, as the basic heading weight is 0.25%, an unreliable estimate will not have an insignificant impact. To illustrate this point, had the 2010 index remained unchanged at 253.55, the Geneva PAI would have been 0.43% higher. This example (albeit simplistic) illustrates how a low number of price observations can result in weak averages with questionable representativeness and reliability. This is why internal and external consistency and plausibility tests should be conducted for each sub-index calculated.

3.1.3 Patterns of volatility:

100. Excessive volatility at item level can possibly be cancelled out in aggregation provided the pattern of volatility was not statistically biased in the same direction. In order to assess whether this might have been the case, we examine the change in basic heading level indices between 2010 and 2016. Table 3.4 shows that large reductions in Geneva indices were more common than large increases.

101. An estimated 71% of the basic headings showed decreases between 2010 and 2016. Given relatively higher inflation in New York than Geneva over the period 2010-2016 this is not necessarily surprising. However, the scale of change is beyond what could be accounted for by inflationary differences. For example, 37% of the items had decreases of between 15pp and 50pp, with a further 10% having decreases in excess of that. The largest decrease was 172pp.

102. At the other end of the distribution 13% of items had increases of between 15pp and 50pp while only 3% had increases of 50pp or more. The largest increase recorded was +67pp.

Table 3.4: Percentage point change of basic headings[25]

	No. of basic headings	% of total	% weight
Decrease of 100pp or more	4	5	0.8
Decrease of 50pp to 100pp	4	5	1
Decrease of 15pp to 50pp	29	37	44.1
Decrease of <15pp	18	23	7.6
Increase of <15pp	11	14	6.8
Increase of 15pp to 50pp	10	13	3.3
Increase of 50pp to 100pp	2	3	1.8
Increase of 100pp or more	0	0	0
Total with decrease	55	71	53.4
Total with increase	23	29	11.8
Maximum decrease	172	-	
Maximum increase	67	-	

Source: Authors own calculations from ICSC documents

3.1.4 Aggregation methodology:

103. For both rents and domestic services, the aggregation methodology was noted as a source of statistical bias due to the application of inappropriate formulas. In those examples, the scale of statistical bias involved could be reasonably estimated with available information.

104. However, there is a wider issue of choice of aggregation methodologies which is at least in part related to the issue of volatility in the indices.

105. In addition to the choice of Dutot indices at the lower level (as discussed in the case of domestic services), a variety of other index number formulas are used for different basic headings at different levels of aggregation, the results of which can significantly differ were alternatives employed. The choice of index is not explained or justified and the reasons are not immediately clear. For example, why are geometric weighted aggregates calculated in one instance (in-area excluding housing) and arithmetic weighted ones used in another (e.g. aggregation of the COLI overall). Readers should note that if

[25] Note: Figures presented refer to the basic headings which could be directly matched between 2010 and 2016. There were a small number of changes but these represent only approx. 1.5% of the total weight.

weighted arithmetic averages were used for all higher-level aggregation on a consistent basis the total index for Geneva would have been over 2% higher in May 2017 than is currently estimated[26].

106. The application of different formulas can have a significant impact on results as demonstrated in the cases of rents and domestic services. In some cases, the choices made are between different valid options where preferred options to those used by ICSC can be identified. However, in other cases the choice of index does not seem suitable for the intended purpose.

3.1.5 Summary of key issues regarding number of price observations, weak averages and volatility:

107. The place-to-place surveys suffer from a number of methodological weaknesses which potentially have substantial impact on validity, reliability and volatility of price comparisons. The weaknesses arise from:

- Low number of items included in the basket of goods and services (relative to other similar exercises)
- Low number of prices collected per item
- Relatively high proportions of collected prices excluded
- Aggregation methods which can increase volatility versus known alternatives

108. The significance of this cannot be underestimated. On this occasion, the observed volatility has clearly worked against Geneva, raising the question are there any other factors which could explain this pattern.

109. The ICSC introduced various approved changes in methodologies, specifications and outlets between the 2010 and 2016. In some cases, these explain why changes might have occurred. For example, for furniture the ICSC switched from collecting prices in a variety of furniture shops to focussing on items in IKEA to improve comparability. The result for that basic heading was a fall from 116.86 in 2010 to 81.10 in 2016. However, these methodological changes would not evidently explain the general pattern of significant decrease in Geneva indices relative to New York.

110. In summary, it would appear that the changes between 2010 and 2016 can be partially attributed to an undesirable degree of volatility, arising from various elements of the methodologies and processes. On this occasion, they have apparently lowered indices for Geneva. However, if weaknesses in the methodology and processes are not addressed, this volatility is likely to be a persistent feature, leading to adjustments in pay based on statistical anomalies rather than genuine movements in the relative cost-of-living. Indeed, it could not be ruled out that in future rounds the changes observed in this round could be dramatically reversed.

[26] Authors estimates based on application of weighted arithmetic averaging to the creation of the in area excluding housing component rather than weighted geometric averages.

3.2 Education

111. The current approaches to calculation of the education index yield results with poor representativeness with respect to expenditures incurred on these groups in New York and Geneva. We believe the range of data collected should be expanded and appropriate weighting structures developed to achieve more meaningful comparisons. It is believed that the index for Geneva is statistically biased downwards as a result of the current methodology with a potentially statistically significant impact. Given the 3% weight attached to this expenditure group, the impact of any statistical bias will be important.

112. Currently the ICSC collects information from various private educational institutions in both Geneva and New York. Average prices are then calculated across three sub-headings which broadly speaking equate to primary, middle and high school.

113. Price collection included a range of private institutions in each location but due to balancing between Geneva and New York a number of institutions were excluded. These exclusions were made in an attempt to ensure only tuition fees were included, as not all institutions provided a schedule of fees which allowed this, leading to their exclusion from calculations.

Table 3.5: Educational institutions in the cost of living comparison, 2016

Duty Station	Number collected	Number included
Geneva	6	3
New York	9	3

Source: Authors analysis of ICSC prices database

114. Table 3.5 again highlights issues regarding representativeness of the average prices being calculated. In general terms, the purpose of the cost-of-living comparison is to compare the average cost-of-living of UN staff based in different duty stations. Under any given heading therefore the items included in the basket should represent the costs incurred by UN staff in each duty station. In the case of education (and also medical insurance) we have serious concerns as to whether the current methodology achieves this objective.

115. If we take the example of other major groups (e.g. Food and Non-Alcoholic Beverages) the approach taken is to specify a wide range of items. While not comprehensive for all items people typically purchase, the hope is that the items are sufficiently representative in both locations to provide a meaningful comparison relevant to all spending on Food and Non-Alcoholic Beverages. In the case of Education (and also Medical Insurance) only one item in each group was selected to represent real price level differences. Furthermore, no weighting has been used to measure the different consumption and expenditure patterns of public and private education in both duty stations. It appears clear that the chosen approach presents a falsely low index for Geneva.

116. The primary explanation for this revolves around the fact that including only private tuition fees does not deliver an index which provides a true reflection of the difference in average education costs incurred.

117. Typically, price statistics are developed using the Classification of Individual Consumption by Purpose (COICOP). In that classification, Education is Major Group 10 and only refers to fees for education. Other education related expenditures are included elsewhere in COICOP (such as transport services, books, etc.). However, in the 2016 ICSC case, only private tuition fees are included in the comparison even though the information from the staff expenditure surveys includes breakdowns of education expenses into categories including transportation, etc. and thus end up included in the education weighting.

118. Table 3.7 attempts to illustrate this by showing two theoretical profiles of staff with education related expenses. The only difference in the amounts is that in profile 1 the child attends a private school and thus incurs tuition fees while the child in profile 2 attends public school.

Table 3.7: Education costs illustration

	Profile 1 Staff member with 1 child in private school	Profile 2 Staff member with 1 child in public school
Tuition fees	25,000	-
Other non-tuition fees	500	500
Books	1,000	1,000
Transport costs	650	650
Extracurricular activities	1,500	1,500
Lunches	1,000	1,000
School clothing	500	500
Total	30,150	6,150
% included in Cost of Living Comparison	83%	0%

Source: Authors analysis of ICSC prices database

119. The obvious question this raises is whether a system which only compares private tuition fees adequately represents differences in the cost of living associated with education. Clearly, even in the case of the person who has a child in private school, this is not the case as a substantial part of the expenses are excluded. However, for the staff member with a child in public school, the comparison simply does not reflect their situation at all even though they do have substantial education related costs.

120. In comparing locations this creates various difficulties:

- The differences in tuition fees between two duty stations may not represent differences in costs for the other elements. In other words, the coverage of the index may be substantially mismatched from the coverage of the expenditure weight.
- If there are different patterns between duty stations in participation in private and public education, a comparison based on private tuition fees only misrepresents the 'average situation' faced by staff (which is more or less shown in the returns to the staff expenditure survey).

121. One possible way to assess whether such an imbalance may exist is to consider the weights for education against the price index for Education (see table 3.8).

Table 3.8: Weights and Price Index for New York

	USD Weight	% Weight	Price Index
Geneva	382.42	2.98%	65.68
New York	306.65	2.71%	

Source: ICSC (ACPAQ/39/R.4 and information provided by correspondence)

122. The price index indicates that average private tuition fees (and thus the entire Education index) are close to 35% lower in Geneva than New York. However, by contrast staff in Geneva spend a relatively higher proportion of their salary on education (approximately 25% more based on the USD weight). For many headings, this may not be contradictory but in the case of education this disparity is not easily explained as each child only receives one education. For both things to be true it would suggest that the demographic profile between Geneva and New York should be very substantially different, with staff in Geneva having, on average, higher numbers of children in education than staff in New York. While possible, the extent of the difference would have to be very substantial (possibly 2:1) to make the relationship between the weights and price index consistent. Assuming this substantial disparity does not exist suggests that the price index is statistically biased low with respect to an index truly reflecting average expenditure on education.

123. One plausible real-world explanation for this is that for reasons of language, public schools are more accessible in New York than Geneva. For that reason, staff in Geneva are more likely to have to register their children in private fee-paying schools. If this is true then it would be clear that the price index of 65.68 does not represent a true comparison of the average situation of staff in both duty stations.

124. Developing an alternative approach would require a study of types of education expenses incurred along with the profile of attendance at different school institutions. However, a substantial amount of information already exists which is relevant to this topic. Firstly, ICSC has the results of the staff expenditure surveys which details the expenditures on different items. Secondly the UN agencies have information from the education grant administration systems. Between them these sources can provide a base to establish a reasonable weighting approach between different items of expenditure. In

addition, they should yield information on attendance patterns at different institutions which can in turn be considered as a source of weighting. Calculating an index composed of more than one education related item would not be particularly complex. In doing so, care should be taken to reflect differences in weights in the two locations to ensure the comparison is technically robust but feasible aggregation methodologies for this are known. An update of this nature could also facilitate inclusion of more institutions in the comparison and therefore reduce the risk of volatility in the index arising from weak averages. Again, all of this should be technically feasible and should generate a more realistic index which almost certainly would be higher than 65.68.

125. In summary, the current education index is based only on one element of education related expenses. In this way, it assumes comparisons of private tuition fees represent a reasonable comparison of education related costs between Geneva and New York. There are many reasons to believe this is not reliable for this purpose. Study should be undertaken to identify an approach which is more comprehensive, representative and reliable. This should be expected to yield an increase in the index between Geneva and New York from the current very low level of 65.68.

3.2.1 Other points related to Education:

126. Two additional points are worthy of note with respect to Education:

- As discussed earlier, during the mission in New York an error was discovered which resulted in the exclusion of valid prices from the calculation of the average price for Geneva for middle school. The secretariat acknowledged this error. The impact of this error was relatively low but should nonetheless be corrected. As also mentioned earlier, of possibly greater concern is the need to have a transparent and easy to implement revisions policy in the event of errors being discovered or other revisions becoming necessary. The impression of the authors is that the current process makes revisions difficult in various ways (e.g. needing to seek Commission approval to correct an error, the IT system being inflexible, etc.). This is an area requiring attention to build confidence in the system.
- A detailed review of the tuition fees included in the comparison indicates some possible incomparability. For example, for some of the institutions included in New York, it seems that additional, non-tuition costs are included for lower grades and it has not been possible to exclude them based on data entered in the database. This highlights a further general difficulty with the pricing approach. By restricting price collection to situations where only tuition fees are identified and only allowing institutions who cover lower, middle and high school, it limits the possible inclusions to very few. This should be reviewed to allow a wider range of information to be considered to avoid having very few price observations driving the index.

3.3 Medical Insurance

127. Medical insurance suffers from a similar problem to Education. Comparison is made by estimating the average cost of medical insurance (available through the UN organisation where the staff member works) in both Geneva and New York. Different premiums from respective policies are weighted together based on staff numbers in each agency. Overall averages for Geneva and New York are compared to generate the index which was 57.24 as of 2016 (i.e. 43% lower in Geneva). To facilitate

making a like-for-like comparison a standard staff member profile is used to estimate the premium payable (P4 Step VI, with spouse and 2 children).

128. A very important element is missing from this comparison, namely, the comparability of the insurance policies. No attempt has been made to standardise the basket priced for each insurance policy. This is important since policies in one location may have greater or lesser coverage than in another, which may generate additional expenditures, either on supplementary insurance or health-related goods and services.

129. The level of coverage provided by medical insurance policies available to the New York agencies appears to be generally higher than those included in the comparison on the Geneva side. Basic policies in Geneva typically require staff to pay an additional supplement to cover 20% of the medical costs incurred while in New York a 100% reimbursement is typically received. As a result, many but not all staff in Geneva purchase supplementary insurance. The cost of this supplementary insurance varies depending on the agency involved, but it can be up to 60% of the base insurance cost and as a minimum tends to cost at least 40% of the base insurance cost.

130. If we temporarily assume that this 40% was added to the index this would increase the index from 57.24 to 80.14. Given the 4.89% weight attached this would increase the entire Geneva PAI by over 1%. Even if we assumed only half of staff took the supplementary insurance the impact would conservatively be over 0.5%.

131. The ICSC in its Questions and Answers document of May 2017[27] explains that the exclusion of supplementary insurance is valid because: "*The comparison of medical insurance costs is a cost comparison, not a price comparison. In other words, this component of the PAI is designed to compare the staff member's out-of-pocket cost for medical insurance coverage at the duty station, with that of his/her counterpart in New York, without accounting for the characteristics of the plans."*

132. This answer does not adequately explain why the chosen alternative should be to include only the two base insurances available. Indeed, if it intended to achieve a cost comparison, then it should include all medical insurance costs. We can note that supplementary medical insurance costs were captured in the staff expenditure surveys. It would have been feasible to calculate an alternative index reflecting average out-of-pocket expenses (as the ICSC wishes to do) reflecting all medical insurance expenditures.

133. However, any adjustment to the system of medical insurance comparisons should be based on a thorough study of medical insurance costs and policy coverage in each duty station. Notwithstanding methodological approvals previously given, it appears clear that the index in its current construction does not reflect the differences in actual expenditures incurred between Geneva and New York. This could be attempted in various ways and yield higher quality results as long as appropriate aggregation methodology is employed. As with Education, in this case it should be expected that such a review would increase the index for Geneva.

3.4 Out-of-area Expenditure

[27] https://icsc.un.org/resources/cold/survey/Explanation_of_COL_surveys_QA.pdf

134. Out-of-area expenditure is a topic which is complex and has been discussed many times. As with all elements of the cost of living comparison, two elements are key namely the weight and the index.

3.4.1. Out-of-area weight:

135. The approach to estimating the out-of-area weight was adjusted between 2010 and 2016. Up to 2010 a common out-of-area weight was included. Following various discussions at ACPAQ and ICSC, it was decided to change this to duty station specific weights. The weight applied for all duty stations up to 2010 was 22.7%. In 2016 the actual level of expenditure recorded for Geneva was 17.01% of all expenditure. For New York, out-of-area expenditure of 10.24% was recorded.

136. ACPAQ and the Commission have both discussed the appropriate approach for incorporating these weights into the calculations. On this basis, a decision was taken which applied wide bands. The result of these bands was that New York was rounded down slightly to 10% while Geneva was rounded up to 20%.

137. During the discussions that led to the approval of the weight bands, various alternatives were considered including using the actual weight, or rounding to bands of 5%. For the 2016 round, these alternatives would yield more or less no difference for New York, but would yield relatively large differences for Geneva, e.g. in the case of 5% bands being used Geneva would instead be rounded down to 15%.

138. The methodology allows that if a weight is rounded down, the excess is allocated to in-area excluding housing.

139. The impact of these differences is not insignificant. For example, it is provisionally estimated that had the actual weight been used for Geneva it would have increased the Geneva PAI by 0.3%, while if a weight of 15% had applied it would have yielded a PAI 0.5% higher.

140. A key justification given for maintaining wider bands was potential volatility for small duty stations. However, this justification does not appear valid for Group I duty stations. The authors would recommend this issue should be reconsidered so as to narrow the bands or use specific weights in the case of Group I duty stations.

3.4.2 Out-of-area index:

141. The out-of-area index involves taking a dollarized average rate of inflation from 26 countries. The same out-of-area index is used for all duty stations. At June 2016 (the date of the New York price collection), this index was reset to 100 and from thereon updated with respect to the different inflation rates and exchange rates involved.

142. The authors acknowledge the complexity of accurately estimating a true out-of-area price comparator, but the current approach is of unclear representativeness with respect to actual spending of staff in different duty stations. As opportunity allows, this would be a worthwhile subject for future study to consider if a more representative alternative could be defined. Given the large weight generally involved this could have substantial impact on results.

3.5 Hotels

143. In 2016 a new basic heading for hotels was introduced. For each duty station an average price is calculated based on hotels in the duty station itself (i.e. within Geneva or within New York). As with other items, these average prices were based on low numbers of observations, however, there is a clear question about the meaningfulness of the index produced. It is almost certainly true that the majority of spending on hotels or other accommodation is not incurred in the immediate area of the duty station. Alternatives should be considered to develop an updated approach which more adequately represents the expenditure incurred by staff. It should not be assumed that New York is representative of hotel prices in the rest of the USA neither should it be assumed that Geneva hotel prices are representative of hotel prices in the rest of Switzerland. While an alternative index cannot be calculated with information currently available, it can be noted that currently this index has a weight of 1.04% in Geneva and had an index of 81.20 thus having a negative impact on the Geneva index overall.

Summary of Education, Medical Insurance, Out-of-area expenditure and Hotels:

- The current approaches to calculation of indices for Education and Medical insurance yield results of poor representativeness with respect to expenditures incurred on these groups in New York and Geneva. It is considered that the range of data collected should be expanded and appropriate weighting structures developed to achieve more meaningful comparisons. It is believed that the index for Geneva in both cases is statistically biased low as a result of the current methodology.
- The approach to incorporating out-of-area weights into the calculations could be reconsidered, at least for Group I duty stations with either actual expenditures or smaller bands used as a more precise approach.
- The conceptual approach to capturing prices for hotels should be reconsidered to develop something more representative of actual staff expenditures. While possibly of minor impact it reflects a wider issue of establishing a clear conceptual base for price comparison which informs pricing and aggregation choices.

ANNEX I

Section 4: The gap closure measure

144. In 2015 the ICSC meeting agreed upon changes to the existing gap closure measures. Specifically, it was agreed to remove the 5% 'cushion' whereby in case of decreases in the pay index of more than 5%, 5% would be added back to reduce the impact on staff.

145. One of the arguments put forward for the removal of the measure was to avoid distortions due to non-application of the actual results of surveys in full. However, the current system introduces other clear distortions which should be noted.

146. To illustrate this, we can view the example presented in Table 4.1 below:

Table 4.1: Pay Index change illustration

	Duty Station 1	Duty Station 2
Pay Index Pre-Survey	165	162
PAI after Survey	155	154
% Difference - Pay Index and new PAI	-6.10%	-4.90%
Change in Pay Index	-6.10%	-
Pay Index after Survey	155	162

147. This example shows that despite having a higher measured index both before and after the surveys, duty station 1 would end up with a pay index 7 percentage points or 4.3% less than duty station 2. This arises because one element of the gap reduction measure was left intact, namely that no change in pay index would be implemented where the reduction was less than 5%, while a reduction of more than 5% would lead to full implementation. While perhaps simplistic the outcome indicated above is feasible and clearly undesirable.

148. With that said, we must consider why there is a need for a gap closure measure. A high level of quality is needed for statistics used for sensitive purposes, such as pay adjustment. However, all statistics will be subject to a margin of error. This is evidently the case for spatial price comparisons such as the place-to-place surveys of the ICSC.

149. As discussed elsewhere in this this report, beyond the clear errors highlighted, there are many areas of uncertainty which raise questions about the reliability of the statistics generated. This includes cases where statistically biased results appear to have been generated plus elements of methodology and practices which create undesirable and unjustifiable volatility in estimated price comparisons. It is the authors' view that with respect to the 2016 round, this statistical bias has, on balance, worked against Geneva. In fact, it is entirely possible that the full calculated decrease in the pay index could be a result of statistically biased results generated from multiple elements of the calculation.

150. However, in the presence of some weak methodologies and practices, statistical bias could work in different directions over time creating a relatively high level of ongoing uncertainty. The gap closure measure is one feasible means to avoid a significant negative impact on staff from results which have a margin of error, which cannot easily be estimated.

151. This uncertainty and margin of error make the reinstatement of gap closure a high priority in order to allow for the fact that, regardless of any improvements made, a range of error will continue to exist and does not offer a stable base upon which to cut staff pay in a very precise manner.

152. Another way to express this is that the current system would see a 5.1% pay decrease fully implemented while a 4.9% decrease would lead to no change. It will never be the case that the statistics generated from the place-to-place surveys can be sufficiently precise to justify this difference in treatment. In fact, it could be argued that the margin of error with current methodologies and processes could be far in excess of 5%, however, in recognition of the precedent of a 5% gap closure measure, this is a reasonable minimum starting point to protect staff from the worst effects of full implementation of uncertain results.

153. In summary, it is clear that a high standard of quality is required of statistics used to adjust pay. The removal of the gap closure measure in 2015 made it effectively impossible for the statistics generated to meet the level of quality required to justify their usage for the purpose in question. Reinstating the gap closure measure creates an allowable margin of error which can be achieved if the right methodologies and processes are applied.

ANNEX I

Section 5 - Conclusions and recommendations

154. The ICSC Cost of Living comparison programme is very complex. From our brief review of the ICSC data, we are satisfied that the price surveys, the family expenditure surveys and subsequent aggregation were conducted professionally. Many of the issues faced and solutions chosen were in line with what could be expected for statistics of this nature. Further the broader system has some very strong features such as independent review, designed to increase confidence in the results. Nevertheless, since the publication of the results, they have failed a number of basic plausibility tests, which directly led to the review team investigation, e.g. seeming implausibility of the rents index, etc.

155. The investigation conducted by the review was necessarily targeted owing to the very brief time available. Thus, as stressed in the introduction, this report cannot be considered a thorough review of ICSC methodology and implementation of the 'approved' methodology. Rather, this review was a brief investigation into a number of aspects of the ICSC results that seemed to have a major impact on the change calculated in the case of Geneva. A summary of facts found during our two-and-a-half-day review of ICSC microdata and aggregations are outlined in this report.

156. Reviewing the 'approved' methodology is in itself somewhat problematic, as the metadata and methodological descriptions are not fully up-to-date, so it is not always clear what is approved and what is not. For example, the ICSC Post Adjustment Booklet and other documents describe how the ICSC uses a modified Walsh index as part of its compilation process. But during discussions with the ICSC, they noted that do not in fact use Walsh indices and this is an outdated reference (see Appendix 3 for more details). The references appear to relate to the use of common expenditure weights when this approach was adopted in the early 1990's. However, even at that time it was not accurate to refer to this as a Walsh index regardless of any modification. It certainly is not correct for the 2016 round. Furthermore, while a large volume of documentation is available, the review team often found it difficult to fully understand the approach taken and ultimately found inconsistencies between reported practices and those in place. The 'approved' methodology only defines and describes practices at a high level. There are no defined or approved methodologies for a range of important processing steps, such as, outlier detection and treatment. Thus, there is considerable room for interpretation as to what 'approved' methodology actually means. From a reviewer perspective, there were also important omissions - for example, full clarity on how 'out-of-area' bands were selected. In the face of this, it is important to note that based on the existing ICSC documentation, no person outside the secretariat (either ACPAQ or others) could possibly have identified the issues presented in this report. The calculation errors arising from aggregation methods, as one example, would not have been easily understood using just the data from ACPAQ/39/R4, Annex V without recreating a complex range of calculations. As a result of this, the review team initially found it difficult, if not impossible to replicate the calculations produced in the ICSC reports, even with full access to all input data.

157. During the course of this brief review, we have identified a number of problematic issues that in aggregate compromise the overall quality and robustness of the ICSC results. We have identified a number of basic but fundamental calculation and aggregation errors that must be corrected before these results can be considered *'fit for purpose'* or sufficiently robust to support decision making.

158. As noted in Section 2, problems were identified with both the formula used to calculate the rent index and the input rent data. Correcting these errors has the combined impact of increasing the rents index for Geneva by up to 20% and increasing the PAI by approximately 4%. Incorporating these corrections alone in to the published PAI for May, decreases the reduction in the pay index for Geneva from 7.7% to less than 4%, well below the 5% threshold for any reduction in the pay index to be implemented. Further, identified issues with domestic services and education should also be immediately corrected, which would lower the level of the reduction even further.

159. Section 3 of this report also identifies a number of other serious issues that while not issues of implementation of 'approved' methodology per-se, still raise disquieting questions regarding the overall quality of the ICSC results and should be addressed urgently. Issues of concern include: weak averages and excessive volatility in basic heading indices, inappropriate use of index formulas; inappropriate approaches to calculating the indices for education and medical insurance and the need to reconsider the weight bands used for out-of-area expenditure. We also have concerns about outlier detection and editing procedures for some sub-indices, and more generally, the distribution of effort between goods and services prices. On balance, it is the view of the authors than on this occasion these issues have resulted in a low-biased index for Geneva.

160. The review team examine in the Appendices some of the factors that determined the differences between the formula/weighting systems used in practice and those that could be justified by standards and practice. In many cases the differences between these will be erratic and volatile. The implication is that the reported figures from the calculations cannot be relied upon and, moreover, if the present practice is continued into further rounds, there is a risk to that quite erratic results may arise.

161. It is the authors' view that resolving these issues is a matter of urgency. They should not however be taken as an indication that the level of quality needed is impossible to achieve. This is absolutely not the case. Short term actions can be taken to allow updated results for the 2016 to be used. Improvements in methodologies and processes are completely feasible and through this report, we have tried to identify areas where we believe the immediate priorities lie. Ongoing review is part of the ICSC process so actions to improve results are already an important part of the process.

162. As discussed in section 4, it is considered essential to reintroduce the 5% gap closure measure removed in 2015 in order to make the level of quality required for these statistics attainable and to give some space to undertake a thorough review and make further changes.

163. A number of recommendations are listed below for consideration. It is the hope of the authors that the implementation of these actions can serve to improve the quality of the statistics and the confidence in them of all interested parties.

5.1 Recommendations

a. Identified errors should be corrected immediately to bring the Geneva PAI to a basic level of feasibility for implementation and ongoing use.
b. The 5% gap closure measure should be reinstated to also cover the 2016 round and future rounds. In our professional view, it was unwise to remove this, as it implies absolute

confidence in the published estimates. Our review has demonstrated that such absolute confidence is not feasible and will still not be feasible even if the methodologies and processes are improved.

c. A comprehensive review of ICSC methodology and implementing systems should be undertaken. The commission should consider initiating such a review. It can be noted that ACPAQ already provides ongoing support for methodological reviews so ACPAQ can be heavily involved or other independent experts could be engaged if desired. ACPAQ should also review the approach for 'approving' methodology as it is clear that additional information and accurate methodological descriptions are required to effectively fulfil its mandate. A review might sensibly begin with establishing a clear conceptual framework for price comparisons to be undertaken and using that to inform price collection and aggregation approaches. The scope of this review should also include the ongoing PAI calculations that are in some part influenced by the issues identified above.
d. A comprehensive review of all ICSC metadata published should be undertaken. While documentation provided is quite comprehensive our informal review identified a number of statements that upon investigation did not stand up to scrutiny and were in fact incorrect or misleading. This is also needed to allow any review mechanism (ACPAQ or other) to adequately perform the task of approving the methodology and its application.
e. A set of statistical quality standards and a code of practice should be adopted and published by the ICSC. These standards and codes should set out in clear terms the quality dimensions appropriate to the objectives of the ICSC and what standard of quality defines *'fit for purpose'*. The ICSC should also clarify whether their statistical estimates are 'official' and whether they are conducted under the auspices of the UN Fundamental Principles of Official Statistics and the UN Principles Governing International Statistical Activities.
f. A clear revisions policy should be published. This policy should clarify the mechanism for implementing corrections in cases where errors have been identified, including a statement on how quickly corrections will be made, whether there is a clear cut-off period (and whether this varies by significance), how this will impact on PAI's and whether the associated metadata will be published.
g. Each round of the ICSC should include a comprehensive peer review by expert price statisticians involving, as necessary, access to microdata once results have been estimated. While different elements of review are currently incorporated they typically do not give reviewers the data needed to fully assess the results calculated either due to timing or data provided.
h. In as far as is possible, all aggregate data (to the lowest level possible without disclosing individual identities) and metadata should be published.
i. Given the escalating importance of services on day-to-day consumption patterns and the consequent increasing complexity of price statistics, the resources allocated to the ICSC should be reassessed. The allocation of existing resources should also be carefully reassessed to ensure that sufficient attention is placed on sub-headings that are driving price and consumption changes. There is a classic 80-20 problem, where 80% of the effort is focusing on items that account for 20% of the impact on change. Considerably more effort should be dedicated to difficult to measure headings such as rent, education,

domestic services, medical insurance and resolving conceptual problems with regard to 'out-of-area' expenditure.

j. The Human Resources Network of the United Nations agencies should consider allocating resources to engage more actively with ICSC and ACPAQ processes on an ongoing basis. This should not be left to staff organisations alone. All UN agencies have a responsibility to their staff to ensure that their pay and conditions are taken seriously. This, however, does not detract from the primary responsibility of the ICSC to ensure the statistics are fit for purpose.

5.2 Note on other duty stations:

164. While the focus of the current review was necessarily on the results for Geneva for 2016, the issues raised are of potential relevance to all duty stations. In particular the issues identified with incorrect aggregation approaches and inconsistent editing processes would have a very direct impact on calculations for other duty stations. The scale of impact of these issues would only become apparent with case by case reviews of the results. It should not be assumed that the identification of a negative statistical bias in the results for Geneva would imply a negative statistical bias for all duty stations. However, the review proposed should build a system which generates higher quality results for all duty stations.

ANNEX I

References

European Commission (2011). *European Statistics Code of Practice for the National and Community Statistical Authorities*, Adopted by the European Statistical System Committee 28th September 2011.

Eurostat/OECD (2012). *Eurostat-OECD Methodological Manual on Purchasing Power Parities*, Eurostat Methodologies and Working Papers, 12th Edition. Available from: https://www.oecd.org/std/prices-ppp/PPP%20manual%20revised%202012.pdf

Expert Group on National Quality Assurance Frameworks (2012): *Guidelines for the template for the National Quality Assurance Framework (NQAF)*

ICSC (2011). 'The Post Adjustment System: What it is, how it works'. New York, February 2011.

ICSC (2017). 'Report on the implementation of the methodology approved by the Commission for the cost-of-living surveys at headquarters duty stations - note by the secretariat of the International Civil Service Commission'. ICSC/ACPAQ/39/R.2, 20 February 2017.

ICSC (2017). *Cost-of-living survey in Geneva*, ACPAQ/39/R4, Annex V

OECD (2011). *Quality Framework and Guidelines for OECD Statistical Activities*, Doc. STD/QFS (2011)1

EUROSTAT/ISRP (2017): *Estate agencies rent surveys: Guidelines for Surveyors, 2017 Version*

United Nations Statistics Commission (2014). *Principles Governing International Statistical Activities*. Available at: https://unstats.un.org/unsd/accsub-public/principles_stat_activities.htm

ILO/IMF/OECD/UNECE/Eurostat/The World Bank (2004), *Consumer Price Index Manual: Theory and Practice*, Geneva: International Labour Office. Available at http://www.ilo.org/wcmsp5/groups/public/---dgreports/---stat/documents/presentation/wcms_331153.pdf

Appendix 1

Rental for housing (Basic Heading 28) - Technical annex on aggregation methodologies

1. The focus here is on two issues:

a. The formula used at the higher level: a Fisher index is used to derive a weighted rental value (price) index across the six categories of dwellings to form the PAI for "Rental for housing." The Fisher index is calculated as the geometric average of Laspeyres and Paasche. However, (i) Laspeyres and Paasche are ill-measured and substantially biased since quantity weights instead of expenditure weights are attached to price relatives; and (ii) the (conceptual) justification for a Fisher index as against a Paasche index is not clear. The two indices give quite different results.

b. The size adjustment factor used to adjust New York (NY) apartment prices for being, on average, larger than Geneva properties for three dwelling categories: three-bedroom apartments, semi-detached, and detached houses. The adjustment is crudely applied as a 25 percent mark-up for three categories and none for three others. The mark-up should be traded for one that takes account of the measured change in the size distribution of the properties and an appropriate valuation of worth the market attributes to an extra square metre of dwelling space. Hedonic regressions, as employed by the Swiss Federal Statistics Office for their rent index, are better suited for this purpose.

1.1 Higher-level formulas to aggregate rental value indices for each of the 6 housing categories (by type and size) into an overall index

2. The higher-level indices suffer from the use of quantity (staff) weights.

3. The weights used for the Laspeyres and Paasche (and thus Fisher) housing indices, at the final level of aggregation by type/size of dwelling, are described as "…using the residential patterns (percentage distribution of staff by dwelling type and size) estimated from the staff expenditure survey."[28]

4. For a Laspeyres rental value index the distribution of relative staff (quantity) numbers in New York is wrongly used as weights and attached to the price comparisons between Geneva and New York of the 6 categories of dwellings (and similarly for Paasche and Fisher price indices.):[29]

$$(3)\quad I_{Las}^{gva\rightarrow ny} = \frac{\sum_{k=1}^{6} p_k^{gva} q_k^{ny}}{\sum_{k=1}^{6} p_k^{ny} q_k^{ny}} = \frac{\sum_{k=1}^{6} p_k^{ny} q_k^{ny} \left(\frac{p_k^{gva}}{p_k^{ny}} \right)}{\sum_{k=1}^{6} p_k^{ny} q_k^{ny}} \neq \frac{\sum_{k=1}^{6} q_k^{ny} \left(\frac{p_k^{gva}}{p_k^{ny}} \right)}{\sum_{k=1}^{6} q_k^{ny}}$$

[28] Of immediate note is that the distribution used appears to be of respondents and not adjusted for differential response rated for categories of dwellings to represent the population.

[29] Such an index is referred to in the literature as a "democratic" index, though not used anywhere, to our knowledge, as a headline formula for inflation.

5. The first term is a Laspeyres spatial price index phrased as a fixed (NY) quantity basket, by the six types and sizes of dwellings, priced at GNV rental values in the numerator and at NY rental values in the denominator. It is the ratio of the cost-of-living in GNV compared with NY for consumption of a fixed (NY) profile of the six dwelling types. The formula benefits from having an intuitive fixed-quantity basket interpretation though, as a Fisher index, has a clear justification on axiomatic and economic theoretic grounds.

6. In practice, the Laspeyres index is compiled as *an expenditure share* weighted average of price relatives, the second expression in equation (3), which equates to the fixed quantity basket definition. A price index with relative quantity weights, the third expression in equation (3), does not equate to a fixed basket definition. Results from quantity-weighted and expenditure-share weighted indices are likely to differ. It is an empirical matter as to the extent and direction in which Fisher will be affected.

7. The use of this quantity-weighted index of price changes can neither be justified on axiomatic grounds nor from economic theory, indeed not even as a symmetric fixed quantity basket index.[30] It is simply not referred to in the international price index manuals and standards.

8. The difference between two formulas (quantity-weighted *minus* expenditure-weighted) determined principally by an expenditure-share weighted covariance for the six categories between the NY to GNV rental values and the NY rental value.[31] For higher (lower) NY rental values positively associated with higher (lower) NY to GNV rental ratios, across types and size brackets of dwellings, the quantity-weighted index will exceed the expenditure-share index, the extent of the difference being determined for the large part by the magnitude of the association. For a covariance of zero—no relationship between NY price and NY to GNV price comparison—the two indices will be broadly similar.

9. *A priori*, the direction and extent of statistical bias between quantity- and expenditure-share weighted price relatives are difficult to gauge. Thus, the use of quantity-weights is not only wrong, but also the nature and extent of the error they induce is difficult to predict and thus can be volatile and unpredictable if used for further rounds. The use of quantity-share weights is both biased and unreliable in measurement formula.

10. The Laspeyres price index in equation (3) is an arithmetic mean of GNV to NY prices weighted by NY expenditure-share weights. A Paasche price index in equation (4) is a harmonic mean of GNV to NY prices weighted by GNV expenditure-share weights that, in spite of its naming, is not equal to the quantity-weighted harmonic mean index, the last term of equation (4), employed in this study.

[30] A Walsh index is a superlative index that makes symmetric use of average (geometric mean) of quantity baskets for Geneva and New York. This Walsh index is given in *ff*. blah and is very different from the "Fisher" calculated here and the "modified" referred to below. This "modified" Walsh is applied for aggregating broad groups of indices into an all-components one.

[31] Derived by the Review team using a Bortkiewicz decomposition; details are available from authors.

$$(4)\ I_{Pas}^{gva\rightarrow ny} = \frac{\sum_{k=1}^{6} p_k^{gva} q_k^{gnv}}{\sum_{k=1}^{6} p_k^{ny} q_k^{gnv}} = \left[\frac{\sum_{k=1}^{6} p_k^{gnv} q_k^{gnv} \left(\frac{p_k^{gnv}}{p_k^{ny}} \right)^{-1}}{\sum_{k=1}^{6} p_k^{gnv} q_k^{gnv}} \right]^{-1} \neq \left[\frac{\sum_{k=1}^{6} q_k^{gnv} \left(\frac{p_k^{gva}}{p_k^{ny}} \right)^{-1}}{\sum_{k=1}^{6} q_k^{gnv}} \right]^{-1}$$

1.2 Use of Fisher price index number formula

11. The Laspeyres price index compares the value of NY household consumption expenditure at GNV prices to the value of this consumption at its own prices. Similarly, the Paasche price index compares the value of GNV household consumption expenditure at its own prices to the value at NY prices.

12. There is a conceptual ambiguity. Price levels will be different on aggregate depending on which basket is being considered. Empirically, the difference between Laspeyres and Paasche is much greater for spatial comparisons than for temporal ones.[32]

13. The finding of quite marked differences in the results for higher-level formula, and the use of superlative index in one case, but not in others, has quite broad implications. If a superlative index is appropriate for rental values, then it should also be used in higher-level aggregation for other components, which is not the case. If the use of a superlative index is not deemed conceptually appropriate, then the rental value index is wrong. In either case the effect on the index of reconciling the formula to a common conceptual base will be substantial and is recommended.

14. On multilateral indices: The answer to a cost-of-living adjustment for an individual duty station compared with NY requires bilateral indices between each duty station and New York, not a multilateral system. Multilateral indices place a premium on transitivity, additively and base-country invariance. The focus on bilateral measures ensures that only GNV and NY prices and expenditures figure are used in a in a comparison between Geneva and NY. Nonetheless, transitivity might well be a useful property to remove what might be deemed as inconsistencies in the measures. There seems to be no reference to the use of multilateral indices in the accompanying notes to ICSC's work in this area, though the use of common weights may well achieve much of this.

1.3 The size adjustment factor

[32] Deaton, Angus and Alan Heston, Understanding PPPs and PPP-based National Accounts, *American Economic Journal: Macroeconomics* 2010, 2:4, 1–35

15. A size adjustment factor is used to adjust New York (NY) apartment prices for being, on average, larger than Geneva properties for three dwelling categories: three-bedroom apartments, semi-detached, and detached houses.

16. New York properties are on average larger than Geneva ones. A size-adjustment factor was used to help balance average rent levels by each of the six categories of dwelling. The size-adjustments factors were 1.25, 1.24, and 1.25 respectively for 3-beddroom apartments, semi-detached and detached dwellings. No size adjustments were deemed necessary for 2- and 1-bedroom apartments and for studios. The impact of these calculations can be quite substantial: the ratio of the average rental value in Geneva to that in NY, for example, for semi-detached dwelling was 1.362 – a 36.2 rent differential. However, this rent differential is boosted by a further 25 percent (size adjustment factor of 1.25) raising it to 1.7021 – a 70.21 rent differential. Similar such adjustments of an additional 25 percent are applied to "detached dwellings" and "3-bedroom apartments". There was no adjustment to the remaining categories of dwellings.

17. On recalculating the rental index without the size adjustment, the index for rental housing is reduced: from 92.47 to 80.00. Of course, there is a rationale for the adjustment. It is the magnitude of the reduction that is of concern, especially given rental for housing accounts for 22.6 percent of the weight of the PAI.

Table A1.1: Rental housing indices for Geneva (New York = 100): Quantity weighted price change

	Size adjustment	
	With	Without
Laspeyres	95.9	86.9
Paasche	89.2	73.7
Fisher	92.5	80.0

18. The size adjustments were used in 2016 in lieu of enhancing the sample of NY dwellings to include smaller properties, a method that proved to be highly problematic. Yet the size adjustments used are questionable. They take the form of a brutal, staged progression of no adjustment for three categories—judged to have no sizable discrepancy—to 25 percent for the remaining three categories. Given the impact of the size-adjustment factor on the measured PAI, some refinement to the size-adjustment factor is required.[33]

[33] What is required is an estimate based on data of individual actual transaction rental values and the corresponding size of the properties rented, preferably those by UN staff. The results from the expenditure survey of individual UN staff give, alongside the rental value, data on the size of the dwelling, number of bathrooms, type of property, a coding for location. It should be a relatively simple matter to estimate a semi-logarithmic hedonic regression: for each property type, regress the (log of) rent on size (and other variables) to derive an estimate of the marginal value on an additional square meter. Appropriate adjustments can then be made. Similar such methods are used for the Swiss rental index as a

component of their consumer price index, though estimates for size are undertaken as square foot per room. Matthey, Christophe and Corinne Becker Vermeulen, *Limitations and Impact of Hedonic Adjustments for the Rent Index*, Federal Statistics Office FSO, Switzerland, 2016.

Appendix 2

The Domestic service index

1. The index for domestic service cost is calculated in two parts: (i) at the lower (unweighted) level of aggregation, arithmetic average *monthly* wage rates are calculated both for Geneva and New York for each of three full-time (FT) categories of domestic staff and *hourly* wage rates for four part-time (PT) categories of domestic staff. The three categories of FT work are: "Maid," "Babysitter/child-minder," and "Other." The four categories of PT work are: "Maid," "Babysitter/child-minder," "Cook," and "Other. Lower-level (Dutot) price indices as the ratio of the GNV arithmetic average wage rate to that of NY are calculated for each of the seven categories; and (ii) at the higher (weighted) level of aggregation these component Dutot indices are weighted to provide two sub-aggregates (higher-level) indices, one for FT and one for PT domestic staff and these in turn are weighted to derive an overall aggregate index for domestic service.

19. Full-time staff is defined as those working more than 160 hours and PT workers as those working less than 160 hours per month.

2.1 The lower-level aggregation formulas

20. The lower level aggregation uses a Dutot index, a ratio of the arithmetic mean of Geneva to the arithmetic means of NY wages for each category of service. An alternative (and preferred) Jevons index, a ratio of geometric means, could yield vastly different results, depending for the main on the difference between the variance of prices in GNV relative to NY. As outlined above, the difference between a Dutot and Jevons index is determined by the difference between the variance of wages; if Geneva wage rates are more dispersed than New York wage rates, then Dutot will exceed Jevons, and if less dispersed, Jevons will exceed Dutot. The magnitude of the difference between the formulas will depend upon the magnitude of the difference in the variances.[34] Undue imbalance or mismatches in the sampling of the workers and service-level at which the service is provided will lead to a change in the variances and thus difference in the results of the two indices. The more heterogeneous the samples, and differences in the variance of the samples, the greater the difference between the two index numbers. The comparison of average prices of "Other" workers is unhelpful in this respect.

21. The Jevons index has stronger axiomatic properties than Dutot and is preferred. The nature and extent of the difference between Dutot and Jevons (a bias in Dutot) is determined by the difference between the variances in wage rates between GNV and NY. This difference can in turn be affected by the measurement process itself including differences in response rates, the detection and treatment of outliers, the sampling of prices. The Dutot index cannot be relied on to be have a small unidirectional bias.

2.2 Higher-level aggregation formulas

22. The index number formula used is statistically biased, and substantially so, in three important respects:

[34] Silver, Mick and Saeed Heravi (2007), "Why Elementary Price Index Number Formulas Differ: Price Dispersion and Product Heterogeneity," *Journal of Econometrics,* 140, 2, 874–83.

- Quantity weights rather than expenditure weights are used;
- Even with expenditure weights, the formula takes the form of a Palgrave index, that has a well-established upwards bias;
- The use of the index is at odds with the higher-level formula used elsewhere in this study and needs to be reconciled with the conceptual needs of the work.

23. The weights used for the aggregation of the Dutot indices at both these stages are the (relative) quantities, that is the relative number of workers employed: 16 FT and 1,302 PT maids; 84 FT and 373 PT child-minders; 42 PT cooks; and 20 FT and 102 PT "Other" domestic service workers. The percentage workers in the Table below act as weights and the division by PT and FT workers reflects the quite different weights attributed to each category and thus the usefulness of the stratification, though the FT weighting of 120/1,77, that is 6.8 percent, is relatively small. The weight given to part-time maids (housekeeping) in Geneva eclipses all other categories; a weight of 1,302 out of a total for full-time and part-time domestic services, i.e. 120 plus 1,819= 1,939, two-thirds of the weight.

Table A2.2.1: Types of domestic workers

Domestic Service	Number		Percentage	
	Full-time	Part-time	Full-time	Part-time
Maid	16	1,302	13.3	73.3
Babysitter/childminder	84	373	70.0	21.0
Cook	-	42	-	-
Other	20	102	16.7	5.7

24. The Palgrave price index is used to compile the domestic service index. The Palgrave index is an arithmetic expenditure-share weighted average of the (average) wage rate of GNV relative to the reference base NY. The weights are expenditure shares for GNV. For a temporal index the Palgrave index would be analogous to a current period (GNV) weighted arithmetic average of price relatives between a base (NY) and current period (GNV). It has an upwards bias and does not have an interpretation as a fixed quantity basket index. The Palgrave index is given as:

$$(5)\quad I_{PAL}^{gnv\rightarrow ny} = \frac{\sum_{l=1}^{L} p_i^{gnv} q_i^{gnv} \frac{p_i^{gnv}}{p_i^{ny}}}{\sum_{l=1}^{L} p_i^{gnv} q_i^{gnv}} \neq \frac{\sum_{l=1}^{L} p_i^{gn} q_i^{gnv}}{\sum_{l=1}^{L} p_i^{ny} q_i^{gnv}}$$

25. In theory and practice, when using current period weights a harmonic mean, a Paasche price index number, should be used (6):

$$(6)\ I_P^{gnv\to ny} = \left[\frac{\sum_{l=1}^{L} p_i^{gnv} q_i^{gnv} \left(\frac{p_i^{gnv}}{p_i^{ny}}\right)^{-1}}{\sum_{l=1}^{L} p_i^{gnv} q_i^{gnv}}\right]^{-1} = \frac{\sum_{l=1}^{L} p_i^{gn} q_i^{gnv}}{\sum_{l=1}^{L} p_i^{ny} q_i^{gnv}}$$

26. A Laspeyres index is a fixed New York quantity basket and is calculated as a New York weighted arithmetic mean of price changes:

$$(7)\ I_{Las}^{gnv\to ny} = \frac{\sum_{l=1}^{L} p_i^{ny} q_i^{ny} \frac{p_i^{gnv}}{p_i^{ny}}}{\sum_{l=1}^{L} p_i^{ny} q_i^{ny}} = \frac{\sum_{l=1}^{L} p_i^{gn} q_i^{ny}}{\sum_{l=1}^{L} p_i^{ny} q_i^{ny}}$$

27. The use of expenditure weighs rather than quantity weights can be seen from the table below to have a major impact on the index: using the arithmetic mean: from 83.05 to 96.75. And using a harmonic mean for all workers: from 82.26 (quantity-weighted) to 93.44 (expenditure-share weighted). Not only is the use of quantity weights wrong, but it causes substantial errors in the results.

28. The use of the Palgrave index as opposed to a Paasche has substantive difference: for all workers from 96.75 (Palgrave) to 93.44 (Paasche), a difference of 3.31 percentage points.

Table A2.2.2: Domestic Service PAI - Geneva Weighted

Quantity-weighted	Full-time	Part-time	All
Arithmetic mean	101.4	81.83	83.05
Harmonic mean	97.65	81.24	82.26
Expenditure-weighted			
Arithmetic mean (Palgrave)	99.27	81.4	96.75
Harmonic mean (Paasche)	95.89	80.84	93.44

29. It is possible to explain why some indices in the above Table are bigger than others. It can be shown that a *weighted arithmetic mean* is equal to or greater than the corresponding *weighted geometric mean*, which in turn is equal to or greater than the corresponding *weighted harmonic mean*.

30. As is apparent from the above formulas, Palgrave's index is not a current-period fixed quantity basket price index—an index that measures the price difference of a fixed Geneva quantity basket of goods and services at Geneva prices with the self-same basket at New York prices. The required index is a Paasche index which translates into a Geneva expenditure-weighted average of Geneva:New York price

relatives by using a harmonic mean. The arithmetic mean, that is the Palgrave index, has no such interpretation and this fixed basket interpretation is largely the basis for justifying alternative index number formula numbers by the economic and axiomatic approaches. The Palgrave index is not used in price index compilation by statistical offices or recommended by international standards.

31. Moreover, the Palgrave index has a well-established upwards formula bias that can be substantial. The demonstration of the statistical bias goes back to Fisher (1922).[35] Diewert (2004)[36] shows the Palgrave index will always be greater than or equal to the Geometric-Paasche which in turn will be always be greater than or equal to the Paasche (harmonic mean) index. The differences can be substantial. Using an illustrative data example, Diewert finds temporal indices to increase over a five period by 67.2, 41.5, and 10.0 percent for Palgrave, Geometric-Paasche, and Paasche indices respectively. The upwards bias infers that for the Geneva:New York price comparison estimates are overstated: the Geneva weighted average of Geneva to New York prices are statistically biased upwards. It is less expensive to live in Geneva than the Domestic service index shows.

32. Using relative numbers of workers as weights yields indices of 101.4 and 0.818 for staff working full-time and part-time respectively to arrive at a weighted average of 0.830.

33. Using relative expenditures (staff numbers *x* average cost) weights yields indices of 0.99266 and 0.814 for full-time and part-time respectively to arrive at a weighted average of 0.968.

34. Out of area housing adjustment: the adjustment for differential size between New York and Geneva of rented dwellings is very crude. The brackets of OA weight do not seem to be working. The cut-off figure was 15 percent with an adjustment of 20 percent applied if this exceeded. If it is not exceeded an adjustment of 10 percent is applied. However, for Geneva out-of-area a correction of 20 percent was applied since it rated 17 per cent, and for New York the rating was 9.78 percent but since this was below 15 percent, the adjustment was only 10 percent.

35. Many Geneva respondents of the survey who live in neighbouring France wrongly recorded their out-of-area housing expenditures in the in-area, rather than the out-of-area, section. These were excluded from the calculation of the weights.

[35] Fisher, Irving, *The Making of Index Numbers*: *A Study of Their Varieties, Tests and Reliability*, Riverside Press, Cambridge Mass., 1922 reference to his Formula (9) distinguished as "…very much biased." (page 112).
[36] Diewert, W. Erwin, Chapter 19 in ILO *et al., Consumer Price Index Manual*: *Theory and Practice*, International Labour Office, Geneva, Chapter 19, page 135.

Appendix 3

Cost-of-living indices

Cost-of-living indices differ in two respects from the index number formulas used.

1. First is the use of common weights. This appears to arise from a sampling issue and while in the long-run it is best dealt with by improvements in the non-response and other such sampling issues, in the short-term consideration should be given to using common weights within bilateral comparisons, that is only for GNV and NY.

2. Second, is the calculation as weighted geometric means of Geneva to NY price relatives? This use of weighted geometric, as opposed to weighted arithmetic means, differs from the weighted aggregation of sub-components of the indices as described above for rent and domestic service.

Table: A3.1: Comparison of Weighted Mean Formulas on Geneva Aggregation

	US $ Weight	Index (NY=100)
Rental for housing	2,902.28	92.47
Maintenance and repair of dwelling	16.64	102.44
Water supply and services related to dwelling	14.48	98.59
Electricity, gas and other fuel	124.2	136.61
Other costs	206.15	125.48
Total	3,263.75	95.76
Arithmetic mean	96.31	
Harmonic mean	95.3	
Geometric mean	95.76	
As Published	95.76	

3. The geometric mean is lower than the arithmetic mean. Geneva is cheaper than New York (less than 1) for both measures, but the arithmetic mean finds prices fell by 3.69 percent while the geometric mean finds it to have fallen by 4.24 percent, even cheaper, and the harmonic mean even more so, by 4.70 percent. With aggregation based on the same weights and price indices, the arithmetic mean will always be greater or equal to the geometric mean, which will always be greater or equal to the harmonic mean (Paasche) index of weighted price relatives. Since the geometric mean is applied to all of (in-area (excluding housing), carrying an expenditure weight of 41.46 percent, the choice of formula is not a trivial matter.

4. There appears to be a somewhat arbitrary element to the choice of formula with arithmetic means, rather than geometric means used at the un-weighted lower-levels of aggregation, weighted-geometric used for indices across items, the Fisher used for components of the rental index, albeit with quantity

weights, arithmetic used for components of domestic service (again, albeit with quantity weights, and arithmetic used to bring the main comport indices together: the 2-digit COICOP level cost-of –living indices, the components of "housing, water, electricity, gas, and fuels," and "in-area (excluding housing)," "pension contributions," medical insurance," and "out-of-area expenditure."

5. Weighted geometric means are not without virtues:

- Superlative indices will lie between Laspeyres and Paasche bounds; indeed, the Fisher index is a symmetric weighted average of the two. If a Paasche is smaller than a Laspeyres price index, then a geometric Paasche is likely to lie within it.[37]
- The geometric indices do not translate into fixed quantity basket forms and thus misses out on this user-friendly interpretation.
- If price relatives become very small or large he weighted geometric men can give implausible results.
- These are described as being calculated by using a "modified Walsh" price index formula. A Walsh index is a superlative index that has good properties with respect to additivity. Of the three commonly known superlative formula, the Fisher, Törnqvist, and Walsh indices, the use of Walsh use is unusual, but not problematic since all three formulas have results that are very close.
- They are robust to outliers

Aggregation to a national index

6. The modified Walsh price index number formula is stated as being used in a number of documents for aggregation to a national index. It is given in ICSC/ACPAQ/39/R.12, Page 8 as a simple weighted arithmetic average of the five major component indices: in-area (excluding housing) for goods and services (as a weighted geometric mean of its items), housing, pension contribution, health insurance, and out-of-area price indices. While this may be a description of the methodology it cannot be reconciled with a Walsh index, however modified.[38] A Walsh index is a superlative index that has good axiomatic properties, though not as strong as Fisher. The description of the formula as a "modified" Walsh misattributes these properties to the index actually used and, in any event, the "modification" to yield this aggregation is far from immediately apparent. We were told that the formula while in the ISCS documentation, is no longer used. Our understanding is that the formula in ICSC/ACPAQ/39/R.12, Page 8 is used; it is the description of the formula as a "modified Walsh" that should be abandoned.

[37] Bert M. *Price and Quantity Index Numbers*, Cambridge University Press, 2010, page 72.

[38] The Walsh price index formula in the first expression below makes use of quantities rather than expenditure weights, w_i^{ny} and w_i^{gnv}; expressed in terms of expenditure weights and prices it is given by the next expressions below:

ANNEX I

7. The modified Walsh index is given by:

$$(8)\quad I_{Mod\text{-}Walsh}^{gnv\to ny} = A\prod_{i=1}^{N}\left(\frac{p_i^{gnv}}{p_i^{ny}}\right)^{w_i^{com}} + \sum_{m=1}^{4} B_m I_m^{gnv\to ny}$$

where A is the weight for the GNV in-area (excluding housing) index, a common weight for goods and services and B_1, B_2, B_3, and B_4 and I_1, I_2, I_3, and I_4 are the respective weights and indices for housing (duty-station weight), pension contributions, health insurance (duty-station weight), and out-of-area components.

8. The formula used is a simple weighted arithmetic average of Geneva to New York price indices of the aforementioned five components. However, the weights vary: the weight for the Geneva in-area (excluding housing) index, pension contributions, health insurance and out-of-area components are duty-station weight. At lower levels, common weights are used to disaggregate the weighting within in area (excluding housing). This pragmatic mix of common weights and duty station weights in itself is not in itself problematic.

9. What is problematic is that the formula for the very large part uses a (duty station) Geneva-weighted arithmetic average of price changes between Geneva and New York – a Palgrave index, when it should be a Paasche harmonic mean

10. Should a superlative index be deemed conceptually desirable the in-area (excluding housing) index would be easily developed as a Törnqvist index:

$$(9)\quad I_{Tor}^{gnv\to ny} = \prod_{i=1}^{N}\left(\frac{p_i^{gnv}}{p_i^{ny}}\right)^{\left(w_i^{gnv}+w_i^{ny}\right)/2}$$

$$P_{Walsh}^{ny\to gnv} = \frac{\sum_{i=1}^{n} p_i^{gnv}\sqrt{q_i^{ny}\,q_i^{gnv}}}{\sum_{i=1}^{n} p_i^{ny}\sqrt{q_i^{ny}\,q_i^{gnv}}} = \sum_{i=1}^{n} w_i^{w}\frac{p_i^{gnv}}{p_i^{ny}} \quad \text{where } w_i^{w} = \frac{\sqrt{\left(w_i^{ny}\,w_i^{gnv}\right)/\left(p_i^{gnv}/p_i^{ny}\right)}}{\sum_{i=1}^{n}\sqrt{\left(w_i^{ny}\,w_i^{gnv}\right)/\left(p_i^{gnv}/p_i^{ny}\right)}}$$