

I

(Information)

COURT OF AUDITORS

SPECIAL REPORT No 10/2006

**on ex post evaluations of Objectives 1 and 3 programmes 1994 to 1999
(Structural Funds), together with the Commission's replies**

(pursuant to Article 248(4), second subparagraph, EC)

(2006/C 302/01)

TABLE OF CONTENTS

	Paragraph	Page
Glossary		3
SUMMARY	I-XI	4
INTRODUCTION	1-9	5
Framework for the evaluation	6-9	6
The basic objectives established for each evaluation	7-9	6
AUDIT SCOPE	10-14	6
EX POST ASSESSMENT OF OBJECTIVE 1 BY DG REGIO	15-80	7
The Court's assessment of the quality of the evaluation	16-66	7
Preliminary remarks	16	7
Evolution of issues confronting the regions since 1994	17-20	8
Appropriateness of strategies adopted	21-28	8
Evaluation of effectiveness	29-40	10
Evaluation of efficiency	41-44	12
Evaluation of the impact of Structural Funds	45-53	12
Evaluation of the management systems	54-66	14
Adequacy of the framework for evaluation	67-75	15
The terms of reference	67-70	15
Adequacy of supervision by the Commission	71-75	16
Relevance of evaluation recommendations for future programming periods	76-80	17
EX POST ASSESSMENT OF ESF OPERATIONS (UNDER OBJECTIVES 1 AND 3) BY DG EMPL	81-113	17
The Court's assessment of the quality of the evaluation	82-113	18
Review of shortcomings in the data and in the analysis	82-87	18
Assessing the impact of ESF operations	88	19

	<i>Paragraph</i>	<i>Page</i>
Evaluation of equal opportunities	89	20
Other aspects examined	90-93	20
Comments on the overall assessment by the Commission	94-96	20
Other insights from the national reports	97-102	20
Adequacy of the framework for evaluation	103-110	21
Recommendations concluding the evaluation	111-113	22
CONCLUSIONS	114-118	22
RECOMMENDATIONS	119-125	23
ANNEXES:		
Annex I — Breakdown of aid under the Structural Funds		25
Annex II — GDP and the employment rate		26
Annex III — GDP growth and the Structural Funds		29
Annex IV — HERMIN simulations and unemployment		30
The Commission's replies		32

GLOSSARY

ADAPT	Adaptation of the Workforce to Industrial Change (a Community initiative to promote employment and the adaptation of the workforce to industrial change)
ALMP	Active Labour Market Policies
CSF	Community Support Framework
Deadweight	Change observed among direct addressees following the public intervention, or reported by direct addressees as a consequence of the public intervention, that would have occurred, even without the intervention.
Displacement effect	Effect obtained in an eligible area at the cost of another area.
ERDF	European Regional Development Fund
ESF	European Social Fund
FIFG	Financial Instrument for Fisheries Guidance
HERMIN model	Macroeconomic model used to simulate the impact of SF interventions
Inception report	Report containing methodological questions such as the evaluation objectives, the methodological framework, the questionnaires and the organisations to be contacted
Monitoring Committee	Committee to oversee the strategic implementation of the funding programme
OP	Operational programme
RTD	Research and technical development
SFs	Structural Funds
SME	Small and medium-sized enterprises
Soft outcome	Result difficult to measure quantitatively
SPD	Single programming document
Synthesis Report	Final reports of DG REGIO (concerning Objective 1) and of DG EMPL (concerning Objective 3) about the <i>ex post</i> evaluation of operations financed by the ERDF and the ESF
ToR	Terms of reference

SUMMARY

I. The audit concerned the Commission's *ex post* evaluations of Structural Funds interventions in Objective 1 and 3 regions over the 1994 to 1999 programming period totalling in excess of 245 billion euro, including matching private and public sector investments.

II. *Ex post* evaluation seeks to compare actual expenditure outcomes against the original objectives. It is the mechanism whereby insights gained during one programming period are exploited to make future SF expenditure more effective. The *ex post* evaluations were carried out by the Commission which relied on contracted external consultants.

III. The audit objectives were to examine if, at the conclusion of the *ex post* evaluation exercise, the Commission was able to estimate the effects of the interventions relative to the objectives and if the results and recommendations have been helpful for the revision of the programmes from 2000 onwards.

IV. The starting point for the audit was the final Synthesis Report for each fund. These reports summarise the overall assessment of the impact of the funds and capture the elements that should be of relevance when taking decisions concerning subsequent planning periods. The next step in the audit was to assess the extent to which any identified deficiencies could be traced back to weaknesses in the ToR or problems during implementation of the evaluation process.

V. Evaluating Structural Fund expenditure is an intrinsically complex process. However, the audit identified weaknesses in the evaluation process that cannot be attributed merely to such inherent constraints.

VI. The Court identified the following weaknesses in the evaluation of SF interventions in Objective 1 regions:

- there were discrepancies in the underlying data,
- the assessment of the effectiveness of SF expenditure was not robust enough due *inter alia* to limited data on outputs and results collected by Member States,
- the evaluation did not fully succeed in comparing the unit costs of projects receiving assistance under the SFs,
- the evaluation was largely based on a macroeconomic model (HERMIN) which suffered from significant limitations.

VII. The Court found that the evaluation of ESF operations under Objectives 1 and 3:

- was in part based on unreliable or incomplete data,
- reached some conclusions which were not adequately supported by the evidence, thereby shedding doubt on the methodology employed to reach them.

VIII. The Commission does not appear to have anticipated some of the difficulties involved in the evaluations in the terms of reference which it formulated for them. Nor did the Commission fully address the shortcomings in the *ex post* evaluation exercise. The Court's findings indicate that there were weaknesses in the oversight of the evaluations.

IX. On the other hand the evaluations yielded useful insights concerning the management systems for ERDF and ESF-funded interventions.

X. However, while some worthwhile and useful recommendations were made as input for future programming periods, the Commission could not benefit fully from the experiences of the 1994 to 1999 programmes.

XI. The Court recommends that

- the Commission should, for future evaluation exercises, ensure that available data and allocated resources are in line with set objectives, which must be realistic; and that the terms of reference must be established accordingly,
- the Commission should give particular attention to the techniques used for measuring the economic impact of the funds. If a macro-economic model is used it must take proper account of the characteristics of the economies under review. Ideally it should be able to benefit from micro-data generated at the project level,
- proper account should be taken of linkages with other studies relevant to the evaluation of the Structural Funds,
- a number of specific issues might usefully be included in future evaluations, relating for example to private sector contributions, 'financial engineering' measures, assessing project costs, project application and approval processes and internal evaluations by national or regional authorities,
- the Commission's oversight of the evaluation process should be improved *inter alia* through collaboration with research institutes and universities.

INTRODUCTION

1. Since 1996, the Commission has been developing the evaluation of Structural Funds within a single framework. The term 'evaluation' is defined by the Commission as the 'judgement of interventions according to their results, impacts and needs they aim to satisfy' ⁽¹⁾. The main purpose of the evaluation activity is to support decision making by providing information and judgements about the relevance, economy, efficiency, effectiveness, utility and sustainability of the impacts of EU interventions. Such information is crucial for improving existing, or designing new, interventions and for setting political priorities and allocating resources.

2. Evaluations are carried out at various levels and stages in the programming cycle (six years) the most important of which are:

- (a) *ex ante* evaluations, conducted prior to the implementation of an intervention to contribute to its design and cost-effectiveness;

- (b) *interim evaluations*, conducted during implementation to improve the management of the intervention;

- (c) *ex post* evaluations, conducted after the completion of an intervention to assess the results achieved and lessons to be learned.

3. The Court's audit concerned *ex post* evaluations which the Commission carried out in connection with SF interventions in Objective 1 and 3 regions over the 1994 to 1999 programming period. For the purpose of this audit, the scope of *ex post* evaluation is taken as that defined in the MEANS collection ⁽²⁾, namely: 'Ex post evaluation recapitulates and judges the entire programme, particularly its impacts. Its aim is to account for the use of resources and to report on the effectiveness and efficiency of interventions and the extent to which expected effects were achieved. It focuses on factors of success or failure, and on the sustainability of results and impacts. It tries to draw conclusions that can be generalised and applied to other programmes and regions' ⁽³⁾. It is against this background that the *ex post* evaluation exercise has been assessed.

⁽¹⁾ SEC(2000) 1051 of 26 July 2000: Communication to the Commission 'Focus on Results: Strengthening Evaluation of Commission Activities.'

⁽²⁾ The MEANS programme was initiated by the Commission as a response to the demand for developing evaluation methodology. The MEANS Collection consists of six volumes on the evaluation of socio-economic programmes.

⁽³⁾ MEANS Collection, No 1, p. 53.

4. The *ex post* evaluation task is part of a 'process' that also draws on the results of *ex ante* assessment and mid-term review. Furthermore, evaluations are undertaken at various levels both national and thematic and culminate in a Synthesis Report. Together they constitute an important element in the management of the Structural Funds and, in particular, the programming and review of Community assistance. Responsibility for the evaluation process rests with the Commission (in partnership with the Member States). It is up to the Commission to decide which aspects of the evaluation to carry out in-house and which it needs to subcontract. In either case, overall and final responsibility still rests with the Commission.

5. Over the 1994 to 1999 period, the aid scheduled under the Structural Funds ⁽¹⁾ amounted to 136,4 billion euro for 'objectives' and to 14,5 billion euro for 'Community initiatives'. *Annex I* provides a breakdown of the former amount by Member State and objective and of the latter by Community initiative. These funds were supplemented by at least a further 95 billion euro of matching private or public sector investment ⁽²⁾ bringing the total investment to over 245 billion euro.

Framework for the evaluation

6. The *ex post* evaluations were carried out on behalf of and under the direction of the Commission. Following a public tendering procedure, a principal consultant was selected for each specified set of objectives and Community initiatives. In most cases, the consultant was supported by a team of evaluators in the individual Member States ⁽³⁾ whose role was to undertake the necessary evaluations of the respective programmes on the basis of a common methodological framework. Moreover, the evaluation process was monitored by a steering group. The evaluations resulted in national reports prepared by the national evaluators and a Synthesis Report prepared by the contracted consultants for each specified set of objectives. These reports were completed between early 2003 and mid 2004.

The basic objectives established for each evaluation

7. The DGs involved opted for a wide-ranging approach covering a broad list of issues. The overall objectives of the evaluations were to:

- (a) assess the results and impacts of the operations in the Member States vis-à-vis their objectives and impacts on persons and systems;

- (b) analyse the impact and effectiveness of the Community actions on specific structural problems and, as far as possible, to ascertain the Community value added of assistance.

8. To meet these overall objectives, the terms of reference (ToRs) established a number of key questions around the themes of appropriateness of strategy, effectiveness, efficiency, impact, management and implementation systems, Community added value and lessons from current and for future programming. To structure the work, the terms also established a number of specified tasks and identified a certain number of techniques and sources of information.

9. The ToR for each objective, as compiled by the Commission, thus provided the basis for the work of the evaluators. These were complemented by the inception report and the methodology guide for Objective 1, in January 2002. For ESF operations, they were complemented by the methodological progress report, in December 2002, and the evaluation questions, included in each national evaluation report. For this purpose, the evaluators relied on literature reviews, questionnaires, stakeholder interviews, the analysis of case studies and other field research as deemed necessary.

AUDIT SCOPE

10. The general audit objectives were to examine if the evaluations provided the Commission with adequate information to estimate the effects of the interventions relative to the objectives and if the results and recommendations have been helpful for the revision of the programmes for subsequent programming periods.

11. The audit focused on the *ex post* evaluations concerning ERDF interventions under Objective 1 and ESF interventions under Objectives 1 and 3, undertaken respectively by the Directorate-General for Regional Policy (DG REGIO) and the Directorate-General for Employment, Social Affairs and Equal Opportunities (DG EMPL). The relevant Synthesis Reports and a sample of national evaluations were examined corresponding to these selected objectives and instruments ⁽⁴⁾ which, together, accounted for 124 billion euro, that is 82 % of the total scheduled aid under the SFs for the 1994 to 1999 programming period ⁽⁵⁾.

⁽¹⁾ Comprising the European Regional Development Fund (ERDF), the European Social Fund (ESF), the European Agriculture Guidance and Guarantee Fund (EAGGF) and the Financial Instrument for Fisheries Guidance (FIFG).

⁽²⁾ Synthesis Report, p. 53.

⁽³⁾ Hereafter referred to as 'national evaluators'.

⁽⁴⁾ National evaluations examined relate to: Objective 1 Belgium, Greece, Spain, Germany, Italy, Ireland, Portugal, and Objectives 1 and 3 (ESF): Spain, France, Italy, Germany, United Kingdom.

⁽⁵⁾ See paragraph 5 for the overall figures.

12. For each evaluation, the audit sought to answer three questions:

- (a) Was the evaluation of good quality?
- (b) To what extent were identified weaknesses due to deficiencies in the framework and supervision of the evaluation exercise?
- (c) Were the resulting recommendations well-based, useful and well-suited to be taken into account in the following programming periods?

13. In particular, the audit assessed the quality and completeness of the reports in terms of their evaluation of the appropriateness of the policies, and the effectiveness, efficiency and impact of the measures, as well as how the measures impacted on the management systems. It also assessed the adequacy of the Commission's role in the preparation, monitoring and supervision of the evaluations.

14. The first part of this report is dedicated to the *ex post* evaluation of ERDF interventions in Objective 1 regions (paragraphs 15 to 80) while the second part covers the *ex post* evaluation of ESF operations under Objectives 1 and 3 (paragraphs 81 to 113).

EX POST ASSESSMENT OF OBJECTIVE 1 BY DG REGIO

15. Before undertaking the detailed assessment of the quality of the evaluation, it is useful to provide a brief overview of the conclusions of the Commission's evaluation. These are summarised hereunder (see Text box 1).

Text box 1

The overall conclusions of the evaluation by DG REGIO

The Synthesis report recognised that the Structural Funds have had a positive impact on the GDP of the Objective 1 regions and that their overall performance, relative to the EU as a whole, has improved. It considered that the extent of the beneficial impact has been heavily dependent on both institutional capacities and factors such as the structure and openness of the economy and that 'these aspects are clearly crucial in particular to the capacity of the economy to respond to the challenges and opportunities of the process of economic integration'. It noted, however, that given the disparity of countries and contexts and because of limitations of the available monitoring data and lack of clearly defined objectives, it was not possible to aggregate the results to give a reliable picture of the overall Objective 1 achievements.

As an overall assessment of the appropriateness of the strategies adopted, the report concludes that the strategic approach to Objective 1 taken in 1994 must be 'viewed as broadly appropriate in each of the Member States'. In its view: 'The most significant criticism of the strategies was that they did not constitute much more than the sum of their parts. Whilst each was an adequate response to the identified needs in particular fields, there was less indication as to how these different elements contributed towards an overarching development strategy.'

In terms of effectiveness, the report indicated that the analysis relied to a substantial extent on 'qualitative judgements rather than quantified comparisons of results with targets' due to shortcomings in the available data. As a general comment, the report states that: 'Many of the apparently most effective actions have been those of a largely "mainstream character." Funding has largely tended to reinforce existing patterns of economic development related activity, rather than providing a basis for developing new activities to meet identified objectives or stimulating more innovative approaches.'

In terms of impact, the report relied, for most regions, on a macro-modelling approach to assess the impact of the SFs on economic and social cohesion. It recognised the fact that: 'The emerging results inevitably flow to some extent from assumptions made within the modelling process.' The report noted that the scale of their impact, in different cases, reflected the relative importance and allocation of the CSF resources and factors such as initial factor endowments, economic structures, levels of competitiveness and the degree of 'openness' of the economy concerned. It drew general conclusions concerning the impact of the structural funds relative to other external factors and indicated that, while job creation has been substantial, dead-weight and displacement was also likely to be substantial.

The Synthesis Report also included a series of observations concerning management and implementation. Its overall conclusion concerning these aspects was that: 'Management and implementation arrangements, though often characterised by weaknesses, were generally fit for the purpose, and significant benefits were achieved through the partnership approach. Decision-making structures typically failed to establish strong strategic direction. However, important benefits in terms of wider "added value" were achieved through the programming approach and the development of public sector management capacity and practices.'

The Court's assessment of the quality of the evaluation

Preliminary remarks

16. It is important to acknowledge that the broad approach adopted faced some intrinsic constraints, the principal ones being: time limitations, a limited range of quantified objectives and a lack of detailed result/impact indicators. Furthermore, the impact that can be attributed directly to the programmes and related measures was often difficult to isolate since the target regions frequently benefitted from other operations that are instrumental in achieving similar objectives. It was also difficult to identify and measure the indirect effects of the Structural Funds expenditure.

However, the Court considers that certain weaknesses in the evaluation process cannot be attributed merely to inherent constraints but were due to deficiencies in the evaluation process which need to be addressed, particularly so before the *ex post* assessment of the 2000 to 2006 programming period.

Evolution of issues confronting the regions since 1994

17. The Synthesis Report starts with a brief analysis of developments in GDP per capita, unemployment and employment rates in the various regions concerned ⁽¹⁾. The report concludes that: 'Increased levels of GDP per head have generally not been the result of increased employment rates resulting from job creation' ⁽²⁾. Such a conclusion would be a potentially important one for policy makers, suggesting that the main goal of cohesion policy of closing the GDP gap has not been significantly helped by improving labour participation rates ⁽³⁾, a conclusion of major relevance to the SFs.

18. From the information presented in the Synthesis Report, it is difficult to judge the validity of the statement quoted above. There are many regions where GDP per capita went up, as did the participation rates. However, there are some regions where either GDP per capita or the participation rate declined or increased only marginally. The Court has sought to assess the claim quoted in paragraph 17 by applying statistical techniques to the data published in the Synthesis Report. The technical details of this analysis are given in *Annex II* and the essential points are briefly summarised hereunder.

19. The results obtained indicate that changes in employment had a statistically significant impact on real GDP per capita. In particular, a one percentage point increase in the employment rate was associated with a 0,24 percentage point increase in GDP per capita over the period under assessment. This result is inconsistent with the conclusion of the Synthesis Report quoted above. It also suggests that there could indeed be some potential to close the GDP gap by improving labour participation rates — a relevant result from the policy perspective.

20. To explore further the robustness of the above result, the Court carried out an identical analysis of 145 regions for which data was consistently available (i.e. not just Objective 1 regions). The results, also reported in *Annex II*, show very similar results across the Union. These results show that information from basic

data was not sufficiently analysed and indicate an important relationship between increases in GDP and increases in the employment rate which the Synthesis Report did not highlight.

Appropriateness of strategies adopted

21. The second chapter of the Synthesis Report examines the appropriateness of strategies adopted in terms of geographical and programme balance, as well as the balance between the funds. The planned and actual programme expenditures ⁽⁴⁾ of the Structural Funds are analysed by Member State and by type of fund but not by the relevant regions — in contrast to what is done in the report on GDP per capita and employment and unemployment rates in the various Objective 1 regions. This limits the analysis that can be performed, as is noted by the report itself in giving the results of a regression analysis which attempts to relate EU funds per capita by Member State to the GDP per head of population ⁽⁵⁾. Accordingly, the results given in the report are not statistically significant and, consequently, are of limited value.

22. However, the Court applied slightly different statistical techniques, using the same data as in the report. The technical details of this analysis are given in *Annex III* and the essential points are briefly summarised hereunder.

23. The results obtained support the intuitive conclusion that the higher the EU funds provided per capita, the higher is the GDP growth per capita in Member States. This result shows that with a broader approach, relevant assessments can be made even with the unnecessarily limited data base used in this important area under consideration.

24. A central part of the assessment in this section of the Synthesis Report is based on a comparison of various categories of planned expenditure in 1994 and the estimated actual expenditure in 1999. However, the data presented suffers from a number of serious errors or discrepancies which limits the validity of the conclusions which can be drawn. In particular, in the case of five Member States, the figure for actual expenditure given under the 'total' column (in the relevant table published in the Synthesis Report) is different from the sum of the figures which appear under the individual columns that denote EU, public and private

⁽¹⁾ Synthesis Report, p. 61.

⁽²⁾ Synthesis Report, p. 62.

⁽³⁾ Here referring to the active labour force as a percentage of the working age population.

⁽⁴⁾ Synthesis Report, Table 2.1, p. 75, Tables 2.5 and 2.6, p. 88 and 89.

⁽⁵⁾ Synthesis Report, p. 75.

sector expenditure respectively (as they appear in the very same table). These discrepancies are summarised in *Table 1*. The percentage discrepancy, which is shown in the last column of this table, is significant and implies an inadequate quality control in the formulation of the report content and results.

25. Notwithstanding these discrepancies, transforming the data given on a per capita basis ⁽¹⁾ provides some useful observations which in some contexts differ from what was concluded in the Synthesis Report. *Table 2* gives the actual per capita expenditure ranked from highest to lowest for each category.

Table 1
Summary of discrepancies noted in paragraph 24

	Sum of columns ⁽¹⁾	Published total ⁽¹⁾	Difference ⁽¹⁾	Percentage discrepancy
Germany	41 015	48 243	7 228	– 15
Italy	34 804	30 547	– 4 257	14
United Kingdom	3 961	4 803	842	– 18
France	3 339	4 019	680	– 17
Austria	1 116	1 040	– 76	7

⁽¹⁾ Million euro.

Table 2
Actual expenditure ⁽¹⁾

Total Actual/capita		EU per capita		Public Actual/capita		Private Actual/capita	
Austria	3 866	Ireland	1 653	Netherlands	1 822	Austria	2 458
Netherlands	3 438	Greece	1 580	Austria	1 033	Germany	1 299
Germany	2 933	Portugal	1 403	Greece	739	Netherlands	1 039
Ireland	2 926	Spain	1 143	Ireland	703	Belgium	705
Greece	2 716	Germany	884	Portugal	621	Ireland	570
Portugal	2 472	Austria	657	Spain	582	Portugal	448
Spain	1 725	Italy	652	France	485	Greece	396
Belgium	1 587	France	621	Belgium	456	France	205
France	1 579	UK	605	UK	381	UK	174
Italy	1 445	Netherlands	578	Germany	311	Spain	na
United Kingdom	1 407	Belgium	427	Italy	na	Italy	na
Average/capita	2 104		1 033		419		377

⁽¹⁾ The same data as published in the report has been used in the compilation of Table 2. As a result, it can be seen that, for the Member States listed in Table 1, the sum of the three columns in this table do not add up to the total column.

⁽¹⁾ Per capita figures, in this context, have been calculated with reference to the whole national population.

26. The table shows that the total actual funds average was 2 104 euro per person. The EU Structural Funds contributed approximately half of this, while three of the four cohesion Member States managed around 50 % more than the average of 1 033 euro per person. Of particular importance is the contribution of the private sector, especially in the light of the emphasis on long-run supply side impacts on the regions. It appears that some Objective 1 regions in Member States, such as Austria, Germany and the Netherlands, have been particularly successful in combining private investment with the Structural Funds, even surpassing, significantly in the case of Austria and Germany, the average contribution from the EU Structural Funds.

27. However, the factors contributing to a higher participation by the private sector remain unexplored in the report. For example, was this a result of the nature of the projects chosen, the skills of the human resources available or indeed a different kind of assistance that is given to the private sector from the funds? In the Court's view, these aspects merited further examination.

28. The assessment of appropriateness was seldom based on formalised techniques ⁽¹⁾ and the practices followed by the national evaluators were quite different and were generally largely descriptive. The matter of appropriateness concerning the degree to which SF interventions address 'market failures' was seldom the explicit subject of *ex post* evaluation. Similarly, the changes in strategy during the programming period have not been examined thoroughly in terms of their appropriateness.

Evaluation of effectiveness

29. Specific chapters in the national evaluations and in Chapter 3 of the Synthesis Report tackle the effectiveness of Structural Funds, based on the extent to which expected effects have been obtained and objectives have been achieved. Five themes are explored, namely Transport, SMEs, R&D, Education and Training, and Rural Development. Two horizontal themes dealing with the environment and equal opportunities are also examined.

30. The analysis is based on a sample of 70 % of the value of Objective 1 programmes. However, a great deal of emphasis was placed on qualitative appraisals, for example rating community

added value on a 'yes' or 'no' basis. Such appraisals and opinions are often classified too broadly and are not always objectively verifiable. There was no attempt to set more concrete objectives by identifying criteria which could have been useful for this purpose and no additional data was gathered for the purpose of micro analysis. Important considerations for a fuller assessment of effectiveness, such as deadweight and displacement effects, were often not considered, and when these aspects were raised, they were not analysed in sufficient detail.

31. Although there are a number of sections which have valid and useful observations, in general the Court finds this crucial chapter ⁽²⁾ in the Synthesis Report lacking robustness, with missing data and often weak conclusions and an insufficient or missing assessment of the results of other thematic studies commissioned by DG REGIO for the SFs for the 1994 to 1999 programme.

32. For example, in the *ex post* assessment section on transport, in the tables dealing with actual and target kilometres of roads improved by country, about half the target information is either missing or incomplete, making it difficult to get a robust conclusion on achievements as compared to targets. Yet, the Thematic Evaluation ⁽³⁾ of the Impact of Structural Funds on Transport Infrastructures for the same programme period, which is quoted frequently in the report, provides a more complete and rigorous assessment of the results by Member State.

33. Of particular relevance to the *ex post* assessment is the estimated employment impact of transport programmes that is given in the Thematic Evaluation Report on Transport. The total direct and indirect employment impact is estimated at 2,3 million person years ⁽⁴⁾. Furthermore the economic benefits of the transport programmes given in the study are judged to be significant. In Ireland, the reported economic rate of return for road projects is given as 13 %, while that of Spain is given as 23 % ⁽⁵⁾.

⁽¹⁾ This assessment needs identification and examination within a global framework of the causal links between the needs, the strategic objectives, the objectives of programmes/sub-programmes/measures and the allocation of resources. It needs also in principle a ranking of the needs and unsatisfactory socio-economic situations which the CSF aim to address and a hierarchy of inherent factors to be considered in order to ensure that resources are optimally targeted towards meeting needs.

⁽²⁾ Chapter 3.

⁽³⁾ A thematic evaluation is an evaluation which transversally analyses a particular point (a theme) in the context of several interventions within a single programme or of several programmes implemented in different countries or regions.

⁽⁴⁾ Thematic Evaluation of the Impact of Structural Funds on Transport Infrastructures, Final Report, November 2000, p. 8.

⁽⁵⁾ For example, in the case of Ireland, this result means that for every 100 euro invested in terms of construction and operating costs, the value of journey time savings, vehicle operating cost savings and safety benefits, was 13 euro for each year. Thus, the investment effectively paid for itself in eight years. (Thematic Evaluation of the Impact of Structural Funds on Transport Infrastructures, p. 8).

34. These two relevant conclusions of the Thematic Evaluation Report on Transport are not taken into consideration in the Synthesis Report. Furthermore, the employment impact estimate in the thematic evaluation is inconsistent with that of the *ex post* study. Based on calculations made by the Court, the total employment impact of Structural Funds over the period 1994 to 1999, implied by the study ⁽¹⁾ (using the HERMIN model and the assessment of the other regions) is 1,62 million jobs, which is inconsistent with the estimate of 2,3 million jobs in the thematic report, especially since transport only accounts for approximately 20 % of all Structural Fund expenditure.

35. In contrast, the results of the Thematic Evaluation of the Impact of the Structural Funds on SMEs are used in the Synthesis Report which states: 'The Thematic Evaluation of the Impact of Structural Funds' on SMEs recommends a shift from grant expenditure to financial engineering measures such as seed and venture capital funds, loans, interest rate subsidies in the future, arguing that these methods are more sustainable in the long term ⁽²⁾. This is an important conclusion within the thematic study which the *ex post* assessment does well to bring out ⁽³⁾. However, the report does not develop this issue further despite the important implications that could be drawn for future programmes.

36. The Synthesis Report quotes the Thematic Evaluation on the Impact of Structural Funds on SMEs as having estimated that 'in the absence of Community support, 70 % of investment projects would have either not taken place at all, or been smaller in scale or postponed'. The thematic evaluation 'further estimates that such assistance contributed to creating more than 300 000 additional jobs, even after taking account of "dead-weight" and substitution effects. The *ex post* evaluation of Objective 1 is not able to replicate these figures owing to the lack of comparable and robust data on which to build such estimates' ⁽⁴⁾. Given the significance of the different results and that the two studies were prepared for the same contractor, treated the same subject, dealt with the same period and were published only two or three years apart, a fuller treatment of their differences was warranted.

37. Following the above statement, in a section dealing with an 'overall analysis', the Synthesis Report does give an estimate of the number of jobs created in SMEs as a result of the Structural Funds. 'From the evidence available it appears that Objective 1 supported the development of some 800 000 jobs. This is in terms of gross jobs and does not take into account substitution,

deadweight or multiplier effects. The overall effectiveness in terms of net effects is unknowable' ⁽⁵⁾. This conclusion is not given sufficient attention, especially in the light of the estimate provided in the thematic study on SMEs. Furthermore, the table in the *ex post* study, giving a breakdown of employment created by Member State, provides no details for Spain, so that the estimate given is clearly incomplete ⁽⁶⁾. This aspect merited a more in-depth examination in the *ex post* evaluation.

38. In the 'other topics' which the Synthesis Report covers under Effectiveness, there is a useful if brief summary of various pertinent issues under a number of headings. However, one underlying limitation is 'weak comparative data which has forced a qualitative assessment rather than a detailed project-based assessment' ⁽⁷⁾. Some of this qualitative assessment could have been better linked to other thematic studies undertaken, such as the thematic evaluation on RTD. In the ToR, it is stated that: 'A broad scoped literature review will be undertaken in order to identify key issues in the implementation of the Structural Funds' ⁽⁸⁾. However, there are shortcomings in the way in which such implications are reported, and are linked to subsequent findings. In the few cases where there is such integration with the assessment, the benefits on the quality of the report are clearly evident.

39. For example, the *ex post* assessment reports that the thematic evaluation on RTD notes that the focus of activities would be better directed at developing region specific innovation strategies to exploit existing RTD capacity, rather than based on perceived 'Technology Gaps'. The *ex post* study reports that this was so in a number of cases, with results being much more successful when focused on local strengths ⁽⁹⁾. This example clearly highlights the relevance of building on what other reports and studies have demonstrated, as emphasised in the criteria developed in the MEANS programme and its update, which was included as part of the ToR of this evaluation exercise ⁽¹⁰⁾.

⁽¹⁾ See section on impact.

⁽²⁾ Synthesis Report, p. 105.

⁽³⁾ The specific comment made was: 'Whilst this would seem a logical recommendation, the context in which it is applied emerges as an important consideration'. Synthesis Report, p. 105.

⁽⁴⁾ Synthesis Report, p. 108.

⁽⁵⁾ Synthesis Report, p. 130.

⁽⁶⁾ To quote again from the study: 'The most effective mechanism for the creation and maintenance of jobs is argued to be direct support for productive investment, although as this takes no account of dead-weight and substitution effects, its overall effectiveness cannot be judged'. Synthesis Report, p. 130.

⁽⁷⁾ The implications of weak data and some general comments on this will be discussed in a later section.

⁽⁸⁾ ToR, p. 90.

⁽⁹⁾ Synthesis Report, p. 111.

⁽¹⁰⁾ ToR, p. 11.

40. The national evaluations were seldom based on a formalised or quantitative analysis of effects. Furthermore, the national evaluators rarely used the results of quantitative analysis carried out by other parties.

Evaluation of efficiency

41. The aim of the Synthesis Report under this heading was the examination of the efficiency in the implementation of large projects supported in Objective 1 regions. The analysis was based on a geographically and thematically representative sample of approximately 60 large projects ⁽¹⁾. Efficiency is understood to refer to the effects being obtained at reasonable cost. 'The main objective is to establish unit costs, i.e. the cost per output achieved. The analysis should also address the quality of the management of large projects by the final beneficiaries. At EU level, unit costs should be analysed across Objective 1 regions. Significant differences in unit costs across regions should be explained where possible and a range of appropriate costs should be established, which may be used as the basis for benchmarking exercises in the future' ⁽²⁾. These objectives are relevant and form a critical element in the efficient management of projects across the EU.

42. However these objectives were largely not met by the *ex post* assessment. This is admitted in the Synthesis Report, with the explanation that project managers found it virtually impossible to determine the individual component parts of the project and subsequently could not calculate unit costs ⁽³⁾. Such a broad statement implies that there is such a poor state of knowledge on major projects funded by the SFs as to make this objective in the ToR unattainable. As a minimum, it should have been possible to arrive at comparable costs for typical outputs, e.g. the cost per square metre of building a factory, or per kilometre of road built.

43. Some partial results are presented in three tables dealing with road construction in Member States, environment projects in Spain and costs per job created in industrial projects ⁽⁴⁾. However, there is little evaluation or analysis of these results, for example to factor in the impact of different labour costs and particular geographic characteristics in say road construction.

44. The Synthesis Report reached firm conclusions about the timing of projects and respect for budget limits. These were, first, that only one third of the 60 projects reviewed were completed within the originally planned timescale, with over a third a year late. Secondly, approximately two thirds of projects examined ran over budget, with 20 % costing over 30 % more than originally planned. However, the analysis is limited to a few general

observations, providing little or no assessment of the quality of management by the final beneficiaries as required in the ToR ⁽⁵⁾. Furthermore, the relative efficiency of public private partnerships and of the tendering procedures employed were not evaluated.

Evaluation of the impact of SFs

45. The Synthesis Report claims that the primary aim of Structural Funds is to influence the long-run supply potential of the economy ⁽⁶⁾. These supply side effects work through a number of channels that are designed to: (a) increase investment in order to improve physical infrastructure as an input to private sector productive activity; (b) increase human capital, due to investment in training, as an input to private sector productive activity; and (c) channel public funding assistance to the private sector to stimulate investment, thus increasing factor productivity and reducing sectoral costs of production and of capital ⁽⁷⁾.

46. The report goes on to discuss the importance of what are called positive externalities — 'conditions where private firms enjoy the use of additional productive factors at no cost to themselves', thereby improving the productivity and cost competitiveness of the economy. These considerations are relevant to an *ex post* assessment. However, private sector co-financing was excluded from the analysis because of uncertainties about its 'driving mechanisms' ⁽⁸⁾. This limited approach leads to a severe understatement of the impact of certain expenditures in the Structural Funds by excluding one of the major long-term supply side effects. One clear demonstration of this is the example of a development company presented in the Synthesis Report ⁽⁹⁾ whose co financed expenditures have attracted significant additional private sector investment. Instead of focusing on the more directly discernible impact of SF expenditure on private sector investment, the study concentrates on attempting to assess the so-called externalities, with rather limited techniques of estimation as a result of a lack of studies providing measures of elasticities which are relevant to the regions under consideration.

⁽⁵⁾ As noted above.

⁽⁶⁾ Synthesis Report, p. 71.

⁽⁷⁾ Synthesis Report, p. 72.

⁽⁸⁾ 'Inclusion of private sector co-financing is of questionable value for an impact evaluation that is based on the use of a formal macro model and since considerable uncertainty and ambiguity surrounds the driving mechanisms behind the private sector CSF co-finance expenditures, we exclude them from our analysis'. Synthesis Report, p. 71.

⁽⁹⁾ UK-Speke Garston project; in this particular example, additional private sector investment of 223 million GBP was attracted, generating 4 600 jobs in direct employment. Synthesis Report, p. 204.

⁽¹⁾ Greater than 25 million euro for infrastructure and greater than 15 million euro for productive investment projects.

⁽²⁾ ToR, p. 13.

⁽³⁾ Synthesis Report, p. 143.

⁽⁴⁾ Synthesis Report, Tables 4.6, 4.7 and 4.8.

47. To assess the impact of the Structural Funds on economic and social cohesion, a macroeconomic modelling approach based on the HERMIN model was applied to four Member States (Greece, Ireland, Portugal and Spain) ⁽¹⁾ as well as for the regions of the new German *Länder* and Northern Ireland ⁽²⁾. This approach was intended to measure the impact of Objective 1 programmes in terms of the effects on GDP and employment.

48. An alternative assessment scenario would be to use microeconomic data from projects to build up to the total or aggregate impact of the Structural Funds on a region or country basis. However, despite the use of standardised data requirements across Objective 1 regions, there are difficulties in collecting microeconomic data. This results in attempts to measure impacts solely from macroeconomic models, giving rise to many restrictions and assumptions which, for the reasons set out below, can restrict the assessment significantly, thereby limiting the policy implications that can be derived.

49. The principal difficulties identified with the specific application of the HERMIN macromodel used in the *ex post* assessment under review are:

- (a) The structure of the model puts too much emphasis on the manufacturing sector while the economies of most of the regions or countries under consideration have seen greater development in the past 10 years in traded market services, such as tourism and, in some cases, the financial sector. The driving mechanisms of these services are different from manufacturing, which has become more susceptible to the impact of globalisation. As a result, the orientation of the macro model needs to be shifted towards this aspect. The model used did not provide for such a consideration and there is little evidence in the report that this aspect has been recognised as a serious deficiency despite the fact that tourism is so important to the economies of quite a number of the regions concerned.
- (b) The model used is based on annual data starting in 1980. However, due to the fast-changing structure of the regions concerned over the intervening time frame, structural instability of the parameters estimated may be a significant limitation of the macro model-based approach adopted. This aspect was also not taken into account.

⁽¹⁾ The four countries which qualify, in their totality, as Objective 1 regions.

⁽²⁾ In the case of Italy the planned construction of a HERMIN model foreseen in the ToR, was not done due to 'contracting difficulties'. For the remaining regions, which did not have a HERMIN model available, a so-called 'bottom-up' estimation was undertaken.

- (c) Use is made of planned rather than actual expenditures ⁽³⁾ which is more reasonably to be expected in an *ex ante* projection rather than in an *ex post* evaluation.
- (d) Impact estimates are significantly biased downwards due to the exclusion of the private sector co-financing from the model (see also paragraph 46) as well as the flow of private sector investment that results from the Structural Fund interventions. For example, if the SFs provide assistance in the building of a factory, the impact on the region's economy is not limited to the construction of the building but relates also to the permanent jobs it gives rise to as a result of the investment by the private sector that takes place following the construction of the factory.
- (e) The impact that Structural Funds have in the long term on the stock of public infrastructure and human capital (positive externalities) plays a crucial role in assessing the impact of the Structural Funds in the Synthesis Report. The HERMIN model requires externality elasticities to be supplied from other specific studies. However, as these elasticities are not available for any of the relevant regions in the EU, use was made of empirical literature largely based on USA regions which, as noted within the Report itself, 'is somewhat ambiguous about the appropriate magnitude of the externalities' ⁽⁴⁾. As a consequence it is not clear to what extent the estimated impacts are based on judgement or on fact.
- (f) As regards additionality, the HERMIN model simulations compare the situations with and without CSF funding. The scenario without CSF funding assumes that neither the EU funds nor the co-financed amounts are spent. The validity of such an assumption is debatable ⁽⁵⁾.

50. The HERMIN model, in the manner in which it has been applied, thus appears unsuited to the task of estimating the effects of SF interventions in Objective 1 regions.

⁽³⁾ The Report states: 'Ideally we should use the actual *ex post* realised CSF expenditures. But these were not available for every country or region, disaggregated by priority and on an annual basis. In the interests of uniformity, we have used the planned CSF expenditure data as contained in the CSF 94 to 99 treaty documents. While these give a fairly accurate total for expenditure, they do not always give an accurate picture of the *ex post* scheduling of expenditures.' (Synthesis Report, p. 150).

⁽⁴⁾ Synthesis Report, p. 161.

⁽⁵⁾ As a matter of fact, some studies claim that EU funding crowds out national investment such that the growth path of the economy without CSF funding would look better than the picture obtained from the HERMIN simulations.

51. The Court analysed the results of HERMIN simulations concerning changes in the rate of unemployment, against a set of pre-established hypotheses ⁽¹⁾. The outcome is shown in *Annex IV*. The results show significant variations regarding unemployment from those available from other sources.

52. Finally, in many cases, the degree to which the intermediate objectives were attained was not evaluated.

53. As regards the strategic objective of convergence, the only mention in the Synthesis Report is 'that the overall convergence between Member States is progressing but that this is not occurring to the same extent between regions across Europe. Whether the convergence is due to the Structural Funds is a debatable point' ⁽²⁾.

Evaluation of the management systems

54. The management and implementation arrangements adopted for the Objective 1 programmes have a crucial role to play in contributing to the effectiveness, efficiency and impact of the Structural Funds. This aspect was addressed in both the ToR and the Methodology Guide and was taken up in the national evaluations ⁽³⁾. The national evaluation reports described the impact on management systems but also identified situations where the effects on management systems could have been more pronounced.

55. The importance of this topic is also recognised in the Synthesis Report and, in the recommendations section ⁽⁴⁾, no less than 13 out of 19 recommendations deal with management issues.

56. The Synthesis Report provides a number of relevant insights into the management of EU Structural Funds and refers to several conclusions of the Thematic Evaluation of the Partnership Principle that was conducted in 1999. This provides for continuity in assessment, and some evidence in support or otherwise of observations made in the earlier study. However, in a number of areas, the conclusions reached are in conflict with the thematic evaluation.

57. The Synthesis Report provides a summary of the average size of membership of Monitoring Committees (MCs) by Member State. The overall average was 44, but there was a wide variation from a low of 15 to a high of 85. Indeed, the report concludes that the high membership level was regarded as a key constraint on the efficiency of the Monitoring Committees, and that this was

common across all Member States ⁽⁵⁾. In contrast, the thematic study concluded that there was no evidence to indicate a loss in efficiency through extending inclusiveness. Given the significant difference in the conclusions of the different reports on this aspect, it would have been appropriate to analyse this aspect more thoroughly in the Synthesis Report, and to draw more concrete recommendations on the way forward.

58. Another area which was linked to the thematic evaluation concerns the management of different funds. Both evaluations concluded that the funds operated according to different principles and with different financial requirements and that this restricted synergy between the funds. This important observation echoes the views already expressed by the Court ⁽⁶⁾ and underlines the need to strive for greater synergy between the funds.

59. The assessment that is made about the effectiveness of the overall management systems also demonstrates some weaknesses. The Synthesis Report states that: 'The management systems were thought to be adequate by the national evaluations in just under half of the eligible member states'. This is an important result and a detailed assessment of both the success cases and the inadequacies identified could have been expected. Although some points raised are developed, there is an overall lack of clarity in the Report about what was actually found.

60. The Thematic Evaluation of the Partnership Principle in relation to the Structural Funds (1999) found that this principle provided a number of positive advantages to Structural Funds implementation ⁽⁷⁾. The Synthesis Report states that 'Broadly, these advantages were also recognised by the Member States involved in the implementation of the 1994 to 1999 Objective 1 programme, and horizontal partnership working was welcomed as one of the "added value" aspects of European programming' ⁽⁸⁾. This is a relevant result, and although some highlights of various cases are presented, including brief comments on negative experiences or difficulties encountered, the discussion could have been more directed at highlighting improvements that could be made, on the basis of the experiences mentioned, so as to make the partnership principle more effective.

⁽¹⁾ This sensitivity analysis was carried out for eastern Germany, Spain, Greece, Ireland and Portugal.

⁽²⁾ Synthesis Report, p. 183.

⁽³⁾ These demonstrate the positive influence of the SFs on the development of multiannual programming, the application of the programme concept and of objective-based management, the establishment of management structures and procedures and in particular systems of monitoring and indicators, systems of quality assurance in public works, tendering procedures for government aid, the development of evaluation systems, capacity and culture, the project selection procedures and control systems. SFs stimulate the development of more appropriate structures and capacities to handle development projects.

⁽⁴⁾ Synthesis Report, section 9.1.

⁽⁵⁾ Synthesis Report, p. 196.

⁽⁶⁾ Opinion No 2/2005 on the proposal for a Council Regulation laying down general provisions on the European Regional Development Fund, the European Social Fund and the Cohesion Fund (COM(2004) 492 final of 14 July 2004) (OJ C 121, 20.5.2005, p. 18); Special Report No 7/2003 on the implementation of assistance programming for the period 2000 to 2006 within the framework of the Structural Funds, paragraph 90 (OJ C 174, 23.7.2003, p. 22).

⁽⁷⁾ These included: greater effectiveness in programme development and monitoring, more effective project selection, greater legitimacy and transparency in decisions and decision-making processes, greater commitment and ownership of programme outputs, etc. See Thematic Evaluation of the Partnership Principle, p. II.

⁽⁸⁾ Synthesis Report, p. 188.

61. The project application/selection process is identified as one of the two most frequently cited processes that caused administrative problems. A number of examples are given, indicating a lack of clarity in the procedures and an urgent need for the Commission to issue a best-practice benchmark. A table is provided giving the duration of the project selection process by country, although no data is provided for three Member States. The duration of the process for the remaining Member States ranges from two to nine months with an average of five months. The evidence provided implies that there is still work to be done in order to streamline the application process. In this context, no thorough analysis has been conducted concerning the respective merits and/or effectiveness in the choice between the competitive process and the queuing process.

62. Due consideration must also be given to the monitoring and control process. It is claimed, in the Synthesis Report, that the 'Financial control systems appear to have functioned relatively well within the Member States under the Objective 1 1994 to 1999 although some adjustments needed to be made to accommodate the new Financial Regulations in 1997. In the majority of Member States financial control systems were felt to be fit for the purpose and operated relatively efficiently. Financial control appears to have been generally reliable and respected the 5 % rule' ⁽¹⁾. This finding does not coincide with the various reports of the Court ⁽²⁾ which indicate that the 1994 to 1999 Structural Funds programme suffered from significant financial control deficiencies.

63. The other element of the monitoring process that gives rise to administrative problems relates to the non-financial monitoring and evaluation systems. Monitoring comes in for heavy criticism. The Synthesis Report concludes that 'monitoring remained a major weakness despite the emphasis on developing its role since the first programming period. The available data is fragmented, inconsistent, overly focused on physical outputs and often of doubtful reliability' ⁽³⁾. It also states: 'Across all of the Member States there was a lack of quantification to the objectives defined at the start of the programme' ⁽⁴⁾. Of the 10 Member States listed, seven are classified as having a poor level of quantification, one as variable and two as adequate only at the start of the programme. Brief summaries are given for each Member State. However, no further insight is given.

64. It is also stated that the computerised systems set up for monitoring the 1994 to 1999 programme 'were frequently criticised due to their lack of integration across programmes and between funds. This had an impact on the ongoing ability of

managing authorities to achieve an oversight on progress within the programmes, as well as posing a problem for later evaluations' ⁽⁵⁾. These findings must also cast doubt on the validity of the earlier conclusion in the report that the financial management of the programme performed well if monitoring information, which is such an important input into the management process, was deemed to have been 'fragmented, inconsistent and of doubtful reliability' ⁽³⁾.

65. The Synthesis Report also provides some brief comments on so-called 'internal evaluations' ⁽⁶⁾. The following observation is made: 'These internal evaluations were carried out by the ministries or regional authorities, with assistance from universities and external agencies. The degree to which evaluation was carried out may, perhaps, be surprising in the light of the weakness of the monitoring systems. This perhaps is a reflection of the perceived requirement to undertake evaluation activity, although the results did often feed through into programme management' ⁽⁷⁾. In view of this statement, it would seem that it would have been useful to focus more on these internal evaluations, and to build on the results achieved by the relevant ministries or regional authorities. In contrast, the Synthesis Report only gives some very general observations with limited relevance to the overall *ex post* assessment.

66. With regard to the 2000 to 2006 programming period, the Synthesis Report notes that attempts have been made to address the monitoring problem experienced in the 1994 to 1999 period ⁽⁸⁾.

Adequacy of the framework for evaluation

The terms of reference

67. In the light of the various problems identified in the analysis of the Synthesis Report it is necessary to focus specifically on the original terms of reference. These, together with the inception report and the Methodology Guide, suffered from certain weaknesses. In particular:

- (a) They did not address the issue of the lack of quantified objectives. Furthermore, the ToR and the Methodology Guide did not sufficiently address the paucity of adequate indicators.

⁽¹⁾ Synthesis Report, p. 218.

⁽²⁾ Annual Report concerning the financial year 2003, paragraphs 5.22, 5.32(d), 5.33 and 5.34 (OJ C 293, 30.11.2004, p. 162). Annual Report concerning the financial year 2004, paragraph 5.30 (OJ C 301, 30.11.2005, p. 143).

⁽³⁾ Synthesis Report, p. 224.

⁽⁴⁾ Synthesis Report, p. 219.

⁽⁵⁾ Synthesis Report, p. 221.

⁽⁶⁾ Dealing with ongoing evaluations in this context.

⁽⁷⁾ Synthesis Report, p. 222.

⁽⁸⁾ For example: 'There is clear evidence that the partnership arrangements have deepened and widened and that there has been increasing delegation of responsibility for decision making to the programme partnership' (Synthesis Report, p. 246).

- (b) Similarly, they did not require the evaluator to collect additional data on performance indicators. The ToR ⁽¹⁾ did foresee the carrying out of field review 'as necessary, if data is not available in the Member States'. However, this point was not further specified in the Methodology Guide ⁽²⁾.
- (c) Insufficient emphasis was put on the need to assess net effects ⁽³⁾, with the exception of some economic impacts.
- (d) The ToRs under-emphasised the question of effectiveness and deadweight effects in particular. This is regrettable given that deadweight is often considered to be a relevant factor in important areas such as business investment grants, direct aid for employment and vocational training facilitating access to employment.
- (e) Insufficient attention was paid to the analysis of the potential links between strategy appropriateness, effectiveness and impact ⁽⁴⁾. Similarly, there was little attempt to relate the identified objectives, even if not quantified, to some ranking of the needs which the CSF aim to address.
- (f) Except for the application of the HERMIN model and the reference to the critical incidence analysis which was suggested for the 'case studies' ⁽⁵⁾, no other evaluation techniques were proposed within the ToR and inception report.
- (g) With regard to the 'country-specific questions' identified in the ToR ⁽⁶⁾, no clear indication was given concerning the methodology to be used.

68. In addition, the strengthening of economic and social cohesion in the EU and the convergence of the less-developed Member States were not accorded sufficient priority in the ToR even though they are amongst the main objectives of structural measures. The inception report did include some questions about the development models used, the constraints on growth which are to be addressed and the channels through which an impact on the regional economy can be expected ⁽⁷⁾. However, the *ex post* evaluation should ideally have included empirical studies on the convergence between regions and the reduction of disparities as

a result of such factors as infrastructure investment, developments in the education system and in R&D capacity.

69. Furthermore, as an evaluation objective, it would have been appropriate to identify and analyse specific case-studies representing benchmarks for regional development policies, for instance, in the area of clustering or growth hubs.

70. At the launch of the evaluation exercise, the question of appropriateness of the strategy was addressed by means of a list of issues and questions. However, the evaluation guidelines provided by the ToR ⁽⁸⁾ and the Methodology Guide were not sufficiently precise or concrete.

Adequacy of supervision by the Commission

71. The (general) contract between the Commission and the consultants selected to carry out the evaluation stipulated that technical assistance with the evaluation was to be provided under the responsibility and supervision of Commission officials. The quality of the evaluation was to be assessed by the Commission on the basis of the criteria developed under the MEANS programme.

72. The steering group chaired by DG REGIO which included representatives of other DGs was involved in drafting the ToR, approving the methodology and detailed work plan (in the Inception Report) and examining the national reports and the Synthesis Report. A panel of four experts was appointed to assist with their principal objective being that of ensuring that high-quality evaluations were produced. However, their intervention came so late so that the process was unable to deliver significant improvements in evaluation quality.

73. The contractually agreed schedule was not adhered to. The submission of the different reports was delayed, with the result that the final Synthesis Report and the national reports were not submitted to the Commission within the planned 12-month period, i.e. by 12 November 2002, but only in April 2003, and were published in May 2003. It would appear that the timetable was excessively tight from the outset and did not include sufficient time reserves for treating the relatively large number of changes/amendments to the reports requested by experts and representatives in the steering group. This has important consequences for future *ex post* evaluations.

74. The ToR ⁽⁶⁾ required the examination of two specific themes by Member State and a list of such themes was proposed. However none of the national evaluation reports included a specific focus on this aspect. Although individual chapters sometimes provide statistics or a partial exploration of these themes, they have not been subject, as required by the ToR, to specific research.

⁽¹⁾ ToR, p. 16.

⁽²⁾ Methodology Guide, p. 16.

⁽³⁾ 'Net' effects are estimated by subtracting from gross effects any deadweight and displacement/substitution effects.

⁽⁴⁾ Especially the links between the results of the analyses of impact and the analysis of effectiveness and appropriateness of the strategy.

⁽⁵⁾ Inception Report, p. 7.

⁽⁶⁾ ToR, p. 102.

⁽⁷⁾ Inception Report, page 13. Similarly, the Methodology Guide includes key questions as: 'What was the overall impact of the structural funds on supporting regional convergence during the period 1994 to 1999? What were the economic effects of the Structural Funds investments in terms of convergence of income levels?' (p. 24).

⁽⁸⁾ ToR, p. 94.

75. As a general conclusion, the oversight of the evaluation process, by the Commission, needs to be improved. In particular, given that the lack of reliable data and performance indicators and of precise objectives was foreseeable from the outset, the Commission should have ensured that appropriate measures were taken to address this critical constraint. Furthermore, given the tight timetable and the inherent constraints, the evaluation should have been concentrated on essential elements rather than attempting to undertake such a wide overall assessment of the programming period.

Relevance of evaluation recommendations for future programming periods

76. Given the problems identified in the preceding sections of this report, the ability of the *ex post* evaluation process to yield worthwhile recommendations was limited. While the process yielded few recommendations about programme content, it delivered a number of relevant and sensible suggestions on how to improve the management of the SFs and of their implementation ⁽¹⁾.

77. The report recommends that: 'The role of the Monitoring Committees needs to be shifted towards one of strategic direction rather than issues of financial management' ⁽²⁾. However, the Monitoring Committees should be able to do both functions, as required by the regulations.

78. The evaluations have led to recommendations for the programming period 2000 to 2006 which mainly address the questions of management systems. A wide range of themes is covered and these are generally well suited to the specific situation of the country concerned.

79. With respect to the 2000 to 2006 programmes, the key lessons to be drawn from the *ex post* evaluation of the preceding period relate to implementation, in particular with reference to the case study work undertaken to evaluate the effectiveness and efficiency of Structural Fund operations.

80. As regards post 2006, the recommendations proposed within the Synthesis Report are rather limited and of a general nature. However, they are still relevant and, as such, are useful. For example, it is stated that 'More use could be made of independent, ongoing evaluations, providing an objective external input' ⁽³⁾. While this is a sensible recommendation,

it would have been of benefit if it had been adopted before the launch of the *ex post* evaluation exercise under review.

EX POST ASSESSMENT OF ESF OPERATIONS (UNDER OBJECTIVES 1 AND 3) BY DG EMPL

81. Together with the evaluation of Objective 1 regions, under the direction of DG REGIO, another evaluation exercise was carried out, in this case under the direction of DG EMPL, to cover ESF operations under Objectives 1, 3 and 4 and the Community Initiatives, Employment and Adapt. The ToR for the latter *ex post* evaluation anticipated the utilisation of the results obtained from the evaluation of the Objective 1 regions. However, despite organisational arrangements intended to facilitate an effective coordination between the two evaluations, there is little evidence that this was achieved in practice.

Text box 2

The overall conclusions of the evaluation by DG Employment

The evaluation aimed at assessing the impact of ESF operations in the Member States, at the level of individuals and of systems. It also sought to analyse the impact and effectiveness of Community action on specific structural problems and to ascertain the Community value added. In the methodology adopted to address these objectives, evaluators were asked *inter alia* to indicate whether the cases which they were asked to assess provided support for or contradicted a number of hypotheses.

The report noted a number of strengths which can be summarised as follows: Much of the ESF funding was targeted at the long-term unemployed which was deemed appropriate given the persistence of the latter. The improvements in the labour market position for the beneficiaries were commensurate with the resources involved. There were major improvements in employment services and educational and training provisions in several national contexts. The resources used enabled a strengthening of certain policy priorities, such as equal opportunities, adaptation of the workforce and targeting those with disabilities. Through the ESF, there was an increased involvement of social partners and regional authorities. The funds contributed to improvements in capacities to manage and implement labour market interventions at all levels.

However, the report also noted a number of weaknesses. Deployment was largely driven by availability of resources rather than policy and took place in parallel and separate from the relevant policy debates. Interventions were largely concentrated on training and mainly allocated to service providers which is not, in isolation, the most effective means of helping the long-term unemployed. Consequently, it is likely that there was a high degree of deadweight associated with the intervention. Finally, administrative arrangements were perceived as complex while evaluation arrangements were weak and unstructured.

⁽¹⁾ For example, a case was made for: 'Providing much greater discretion to those Member States which have effective domestic monitoring systems to utilise their own monitoring requirements in relation to EU-supported initiatives. This would reduce the costs involved and the data which emerges would probably be more meaningful'. Synthesis Report, p. 249.

⁽²⁾ Synthesis Report, p. 249.

⁽³⁾ Synthesis Report, p. 251.

The Court's assessment of the quality of the evaluation

Review of shortcomings in the data and in the analysis

82. The influence of ESF interventions often takes the form of a contribution or of a potential effect that is intangible and therefore not easily quantifiable⁽¹⁾. Notwithstanding this intrinsic problem, it must still be a prime objective of an evaluation exercise to assess the impact and, in conjunction with it, also the question of effectiveness and efficiency. This is what the *ex post* evaluation exercise was required to do.

83. In seeking to analyse the impact of ESF interventions, the Synthesis Report provides estimations of the number of ESF beneficiaries per country and per objective⁽²⁾. As a complement to the national evaluation reports, the *ex post* evaluation also attempted to provide evidence for a number of pre-established hypotheses⁽³⁾. These considerations are discussed in the following paragraphs.

84. One of the most basic requirements of the *ex post* assessment of ESF operations relates to the number of beneficiaries of the programme over the period 1994 to 1999. The report provides a table which gives the estimated number of beneficiaries as 52 million. However, the additional notes to this table cast serious doubts on the figures presented⁽⁴⁾. The uncertainty about the accuracy of the data detracts from some of the useful and relevant observations resulting in other parts of the assessment.

85. The Synthesis Report⁽³⁾ provides estimates of impacts on individuals from ESF interventions. This information is potentially important in a number of aspects, including the calculation of unit costs of programmes and estimates of beneficiaries still in employment after 12 months. However, there are factors which seriously limit the usefulness of the information, notably the lack of explanation as to how the data has been derived. The report simply states that: 'These ratios applied to generate these estimates have been informed by the individual cases examined and the review of the literature'⁽²⁾. Such a statement cannot be considered as being sufficient and more detailed information on the derivation of the various estimates is necessary to understand better the applicability of these numbers, such as benchmarks of unit costs and success rate of programmes.

⁽¹⁾ For example, increased labour force 'flexibility' as a result of training.

⁽²⁾ Synthesis Report, p. 124.

⁽³⁾ Synthesis Report, p. 125.

⁽⁴⁾ For instance, one can read in these notes that the figures for France, which amount to 17 million, were being 'double-checked' for 'double counting' while the figures for the Netherlands, which amount to 8,7 million, are also being 'double checked with national teams'. Furthermore, it is indicated that data are largely missing for Germany, Ireland and the United Kingdom — although the latter is listed as having 7,9 million beneficiaries and Ireland as having 680 000 beneficiaries.

86. A second unsatisfactory aspect is the analysis accompanying the data. There is a mismatch between what is stated in the assessment and what is found in the accompanying table. For example, the text states that 'in broad terms unit costs were lower for the unemployed and young unemployed, around 6 000 euro per worker in employment, and considerably higher and wide-ranging for those with disabilities and long-term unemployed'⁽²⁾. Yet the table gives the unit costs for training the unemployed as the highest (at 6 000 euro) and the integration of the disabled as one of the lowest (at 4 000 euro). Furthermore, it is common practice to give the range of variation when there is what is claimed to be 'wide variation' between and within programmes. The analysis in this section is of limited value because it relies overly on reducing the assessment to a single number rather than defining a relevant range of values and providing a more comprehensive analysis of these ranges found in the various cases examined.

87. This mismatch between the commentary and the figures given occurs in other parts of Chapter 3 (see Text box 3 for some examples).

Text box 3

For instance, it is stated that: 'The ESF per capita expenditure was highest in Ireland, Portugal and Greece (fully Objective 1 countries) and in Spain⁽¹⁾'. Yet Table 3.1 gives an ESF expenditure per capita per year for Portugal of 2 086 euro and for Germany of 405 euro, while Ireland is given as 12 euro and Greece as 17 euro. Clearly, the figures in the table do not match the discussion in the text. Even more significant is the statement: 'The ESF expenditure on education relative to national expenditure was in all countries lower than the relative expenditure to ALMP. It was highest in Greece and Portugal⁽²⁾'. Yet in Table 3.1, Greece is classified as having no ESF expenditure in the relevant column giving ESF expenditure on education relative to national expenditure.

For example in paragraph 3.2.3⁽³⁾ it is stated that: 'The national expenditure on labour market training as a proportion of GDP ranged from 0,2 % in Luxembourg to 0,58 % in Sweden⁽⁴⁾'. Yet in Table 3.3, the expenditure on training in Denmark is shown as 0,94 % of GDP, almost twice that of Sweden. As another example, consider paragraph 3.2.6⁽⁵⁾, dealing with expenditure on the integration of the disabled. In the text, one finds the following statement: 'The ESF was particularly important in Portugal, Spain, Greece and Austria.' Yet in Table 3.6, Spain and Greece are listed as having no ESF expenditure on the disabled. A further example is found in paragraph 3.2.7 dealing with expenditure on education. It is stated that: 'The ESF contribution was most significant in Greece, Ireland and Portugal⁽⁶⁾'. But Greece is listed in Table 3.7 as having no ESF expenditure on education and training.

Synthesis Report: ⁽¹⁾ p. 78; ⁽²⁾ p. 79; ⁽³⁾ p. 81; ⁽⁴⁾ p. 81; ⁽⁵⁾ p. 84; ⁽⁶⁾ p. 85.

Assessing the impact of ESF operations

88. The evaluators who took part in the study were asked to indicate whether the cases assessed provided support for or contradicted a number of hypotheses that had been formulated on the basis of an initial review of the literature and the findings of the inception phase of the evaluation. The study claims 'strong

supporting evidence for four hypotheses' (see Table 3 ⁽¹⁾). However, the data presented does not support the claim of strong supporting evidence. For example, only seven of the 30 evaluators said there was 'strong supporting evidence' for the first hypothesis. A more objective evaluation of these results would have been warranted.

Table 3

Observations on the hypotheses addressed in the examination of cases of impact on individuals

	Strong supporting evidence	Some supporting evidence	No evidence to support or contradict	Some evidence to contradict	Strong evidence to contradict	No opinion at this stage
1. The level of impact achieved is related to the underlying national and regional market conditions (e.g. results depend upon the economic cycle and level of demand and unemployment).	7	14	8	0	1	0
2. The level of impact is related to the characteristics of the beneficiary groups targeted. (e.g. training measures are more effective for women re-entering the labour market and educated immigrants than for low-skilled workers and youth).	8	18	3	0	0	1
3. The level of impact is related to the type of measure. Some measures show high levels of deadweight and some are effective at a small scale but do not provide general solutions.	5	17	6	1	0	1
4. The level of impact is closely related to aspects of the structure and management of programmes including the quality of analyses that define the precise scope of interventions. (e.g. small targeted programmes may be more effective than large-scale programmes).	3	17	7	1	0	1

⁽¹⁾ Table 3 reproduces part of Table 4.8 in the Synthesis Report.

Evaluation of equal opportunities

89. The terms of reference concerning the impact of ESF 1994 to 1999 also required an evaluation of the extent to which ESF operations provide support to the 'equal opportunities ⁽¹⁾' objective. This aspect is tackled in Chapter 7 of the report. However, it is claimed that: 'A major finding is that gender disaggregated data were not collected systematically for each of the ESF Objectives and Community Initiatives across the EU' ⁽²⁾. This has limited significantly the ability of the study to evaluate this topic. It is difficult to accept this result as a major finding of this report since a preliminary analysis would have revealed this shortcoming, necessitating field work and sample surveys to overcome this restriction, rather than simply concluding that disaggregated data is not available.

Other aspects examined

90. As regards the impact of ESF operations on individuals, the interventions 'combined measures' ⁽³⁾ have not been specifically examined in spite of the substantial percentage of ESF resources having been allocated (18 %) to them. As a result, the conclusions of the Synthesis Report are rather modest: 'Most of the evaluation work points to positive outcomes. However, the evaluation methods tend to focus on identifying the components of the programmes that are most effective rather than judging the performance of combined measures compared with single measure interventions' ⁽⁴⁾.

91. As regards the impact of Active Labour Market Policies (ALMP), especially with regard to Objective 1 countries, the Synthesis Report draws some conclusions without adequately explaining how they were reached.

92. As regards the relative financial contribution of the ESF, the Synthesis Report ⁽⁵⁾ concludes: 'that the combined allocation of national expenditure and ESF resources does not appear to have led to a convergence of the patterns of expenditure between the countries, nor the adoption of norms linking specific economic and labour market conditions to resource allocations' ⁽⁶⁾. It is difficult to see how this conclusion can be sustained given the level of errors in the relevant tables and/or mistakes in the discussion. Similarly, another important conclusion in this part of the report is that: 'These circumstances suggest that not all of the ESF interventions are likely to have been additional. Additionally, in the sense of the ESF co-financed interventions being unlikely to otherwise have taken place, is more likely to have occurred where ESF resources were focused and concentrated on particular types of interventions and where the related national

expenditure and programmes were relatively small and less developed compared with countries with similar labour market conditions.' Here again, it is difficult to see how this conclusion is reached from the evidence presented in this chapter, especially given the mismatch that there is between the analysis and the data given in the tables.

93. With regard to the hypotheses advanced on the impact of ESF on systems, a problem arises similar to that described earlier. Strong supporting evidence is claimed this time for seven hypotheses ⁽⁷⁾. However, an analysis of the data presented also indicates that the conclusion is not warranted.

Comments on the overall assessment by the Commission

94. Giving due weight to the constraints faced by the evaluators, the following critical comments must nevertheless be made with regard to the Commission's own assessment of the *ex post* evaluation which concluded with a satisfactory overall appraisal.

95. The evaluation of the report by DG EMPL is based on eight performance criteria and four possible quality levels, namely: excellent, good, acceptable and unacceptable. None of the applied criteria are rated as excellent or unacceptable. Four criteria are rated as good, namely: meeting needs, relevant scope, defensible design and clear report. The four other criteria considered are ranked as acceptable. These relate to reliable data, sound analysis, credible results and impartial conclusions.

96. It is significant that DG EMPL assessed the latter four criteria as being less satisfactory than the previous four. However given the problems encountered as regards reliability of the data and the inconsistencies of the commentary thereon, as highlighted in this report, the Court considers that the evaluation report was unsatisfactory, especially with regard to reliable data, sound analysis, credible results and impartial conclusions.

Other insights from the national reports

97. Some pertinent comments can also be made with reference to the national evaluation reports, with regard to the evaluation objectives and results:

- (a) The evaluation results in relation to the impact on individuals were often of a qualitative nature; deadweight was not always estimated.

⁽¹⁾ Here refers primarily to gender equality.

⁽²⁾ Synthesis Report, p.173.

⁽³⁾ Combined measures enclose intensive counselling and training, work experience and training and employment incentives and training.

⁽⁴⁾ Synthesis Report, p. 122.

⁽⁵⁾ Paragraph 3.2.8.

⁽⁶⁾ Synthesis Report, p. 86.

⁽⁷⁾ Synthesis Report, p. 144.

- (b) Impact aspects of a more qualitative nature ⁽¹⁾, the so-called soft outcomes, have not been systematically examined.
- (c) Except for integrated or mixed approaches, few general statements concerning the performance of individual measures can be deduced from the evaluations.
- (d) Important factors such as the contribution to sustainable development and to socio-economic cohesion have not been systematically assessed.
- (e) Target groups such as the long-term unemployed and young people seeking to enter the labour force have not systematically been the subject of specific conclusions in the evaluation of the co-financed measures.
- (f) Only in a few cases was it possible to compare expectations with results; the evaluation was generally hampered by the lack of specific initial objectives and indicators and the unavailability of accurate information.
- (g) The complementarity between ESF and other actions has only rarely been taken into consideration in order to explain certain results. The significance of the productivity increase of individual ESF beneficiaries, in terms of the benefits to the firms employing them, was almost never examined directly.

98. The sampled national reports do not always present a thorough analysis of the impact of ESF interventions on national policies. However, they generally do suggest that the ESF had an impact on the development of ALMPs and, in some cases, a concrete, direct effect was stated. However, in some other instances, a different assessment was made, namely that it is difficult to show that the ESF has led to changes in the ALMPs.

99. Each national evaluation report and the Synthesis Report describe in a more or less extended way the impacts on systems as regards employment services; initial education and training; continuing education and training; accreditation systems and mechanisms for anticipating needs. The impacts on systems often concerned a more specific policy context. It was generally difficult to identify the net impacts of ESF operations ⁽²⁾; only in exceptional cases, was the net impact effectively delimited ⁽³⁾.

⁽¹⁾ For example, the achievement of qualifications and the soft outcomes presented in the UK evaluation report (representing 'distance travelled toward the labour market').

⁽²⁾ This was also stated by some evaluators (Spain and Italy).

⁽³⁾ For example, in Spanish cases.

100. Concerning the European Employment Strategy, the evaluation reports highlighted the positive influence of the ESF in some Member States such as Greece, France, Italy, Portugal and Spain. However, it was noted in the case of Germany that the top-down approach associated with the European Employment Strategy had the potential to conflict with the ESF bottom-up approach ⁽⁴⁾. At the more general level, the Synthesis Report concludes that the EES has provided a stronger strategic framework within which to programme ESF interventions ⁽⁵⁾.

101. As regards the Community value added, the national evaluation reports have contributed some relevant results ⁽⁶⁾, although some of the specified questions have not been dealt with in some national reports.

102. The financial implication of policy choices has rarely been addressed in detail. The Synthesis Report includes a calculation of unit costs concerning some measures but this is based on unsatisfactory data. Where cost issues have been addressed, they would seem to highlight potentially serious findings that merit further analysis. For example, for Spain, it is claimed that 'This roughly doubling of resources due to the ESF has led, according to some experts, to a substantial amount of wasteful spending. The larger source of funds has been used to expand operations and not so much to introduce new measures.' Similarly for Italy, it is claimed that there have been relevant transaction costs associated with the ESF, referring to implementation issues and the costs associated with partnership building and new management practices ⁽⁷⁾.

Adequacy of the framework for evaluation

103. The following main evaluation topics were specified in the ToR: the principal characteristics and the accomplishments of the 1994 to 1999 programming; the impact of the ESF vis-à-vis national policies and the effectiveness and impact of ESF operations on individuals and on systems.

⁽⁴⁾ One such example was quoted in the case of Germany.

⁽⁵⁾ In particular, the EES promotes stronger links between EU programme funding mechanisms and other policy instruments, regulations and national programmes including passive labour market policies.

⁽⁶⁾ For example, in the case of Germany: support to the objectives that target disadvantaged sections of the public, sustainable development, equal opportunities, financial leverage effects for the new *Länder*. Similarly, various other topics were highlighted in the other national reports.

⁽⁷⁾ National report, p. 95.

104. A methods report was compiled by the Commission and published a year later to provide a critical review of the methodologies adopted for the assessment of results and of their impacts on individuals and systems. This report was intended to prepare the ground for future evaluations but it is questionable if this objective has been effectively achieved. This report suffers from various weaknesses. Its preparation was not based on a sufficiently thorough enquiry as regards to the approaches applied in assessing the net impact on individuals. As concerns the evaluation of the impact on systems, it does propose a model for the classification of ESF expenditures and of their anticipated associated impacts (capacity, capability and responsiveness) ⁽¹⁾ though some issues remain to be addressed (see page 35 of the report).

105. In the Methods Report, within the context of the 'Methods for policy and strategic evaluation in the ESF field' ⁽²⁾, no concrete recommendations have been presented as regards policy analysis, with the report merely referring to MEANS. Similarly, there was no reference to the approach to be followed for the selection of appropriate indicators at different strategic levels.

106. It should be remarked that the relevant elements for some evaluation tasks are not always clearly specified in the ToR. For example: the description of the tasks as concerns effectiveness is rather abstract, especially with regard to the underlying concepts, the selection of mechanisms for impact measurement and the evaluation techniques to be used. In particular, no concrete indication has been given as regards the methods of generalising, at the EU level, the direct impact on individuals from the results of the activities in the different Member States. The formulation of the tasks has not given due regard to the possible mismatch between quantitative objectives and the data available. Furthermore, no concrete action has been foreseen to evaluate the impact on individuals when the indicators are not concretely established, or when specific statistical evaluations have not been carried out recently. The other methodological documents, compiled subsequent to the ToR, do not contain the desired explanations either.

107. Changes in programming have not always been adequately examined. For example, this subject was not addressed in the studies concerning Northern Ireland, Merseyside and Spain. Even where changes in programming have been indicated, their treatment is far from being optimal. In some cases the evaluator has not provided a clear assessment of the relevant changes and of their appropriateness ⁽³⁾.

108. The Commission did not carry out a pre-evaluation survey or reconnaissance exercise and this has made it more difficult for it to anticipate the problems that have been encountered by the evaluators.

109. There is no clear evidence that the steering group was sufficiently aware of the lack of data available to the national

evaluators, nor the impact that this would have on the quality of the final evaluations or the problems regarding delays. No remedial action was taken by the steering group to address this issue.

110. As was the case for the Objective 1 evaluation, the Commission's oversight of the evaluation process needs to be improved (see paragraph 75).

Recommendations concluding the evaluation

111. For most Member States, the recommendations derived from the 1994 to 1999 *ex post* assessment have mainly dealt with the management systems and procedures. They relate to a variety of issues and are generally well adapted to the specific situations.

112. Fewer recommendations for 2000 to 2006 have concerned the content of the co-financed policies. Where such recommendations have been made, they appear to be of relevance ⁽⁴⁾.

113. The Synthesis Report brought up a list of strengths and weaknesses for each Objective ⁽⁵⁾ and main recommendations for the 2000 to 2006 period and post 2006 ⁽⁶⁾. These statements are rather general although they do appear to be valid in certain cases, e.g. in highlighting potential differences between EU-15 and EU-10 and in the policy response to territorial differences.

CONCLUSIONS

114. The evaluation process is of critical importance in terms of providing information and judgements about the appropriateness, relevance, economy, efficiency, effectiveness and the impacts of EU interventions. By doing so, it can support the decision-making process for setting political priorities and allocating resources. In order for the process to be successful, however, there are certain prerequisites to be respected, namely, clear objectives for policies and programmes, and performance indicators which are produced on a regular basis and which are complete, reliable and relative to the objectives set. In addition, it is necessary to ensure that the resources available to carry out the process are adequate for the purpose, both in terms of time and technical competencies.

⁽⁴⁾ Examples are: the need for a stronger link between the planned objectives and the local socio-economic context (Italy); a better targeting of beneficiaries (France); the expansion of integrated intervention approaches (United Kingdom); to widen the scope of the programmes to more innovative and risky actions (Spain); to strengthen anticipation actions in the training domain (Spain); to pay more attention to the less-favoured regions by putting more weight on variables such as unemployment, long term unemployment and social exclusion (Spain).

⁽⁵⁾ Synthesis Report, p. 188 for Objectives 1 and 3.

⁽⁶⁾ Synthesis Report, p. 198.

⁽¹⁾ Methods Report, p. 24.

⁽²⁾ Methods Report, p. 44.

⁽³⁾ See evaluations concerning France, Spain and Italy.

115. The Court's audit of the *ex post* evaluations under review identified a number of significant shortcomings in the overall approach adopted and in the quality of the assessments made:

- (a) Objectives set during the 1994 to 1999 programming period lacked clarity and coherence.
- (b) While there were few performance indicators available, few measures were taken to overcome this deficiency by, for example, collecting additional quantitative data.
- (c) As a result, an appropriate balance between quantitative and qualitative analysis was not always achieved.
- (d) This situation resulted in conclusions being drawn which were not supported by adequate analysis.
- (e) At the same time, a number of themes which merited being treated in an *ex post* evaluation, such as deadweight, convergence and the proposed themes per Member State, were not always dealt with in the evaluation.

116. With regard to the various themes of the evaluations, the following conclusions can be drawn concerning the Objective 1 evaluation:

- (a) In assessing the appropriateness of the strategies adopted, the analysis suffers from some serious discrepancies in the data, as published in the Synthesis Report, concerning the comparison between planned and estimated actual expenditure. Even allowing for limitations in the data available, a more in-depth analysis could have been attempted to shed more light on private sector investment linked to SF expenditure.
- (b) The examination of effectiveness lacked robustness, *inter alia* due to inadequate or insufficient analysis and a failure to take into account some important and relevant conclusions in thematic studies.
- (c) The examination of efficiency was confined to some general observations, while the evaluation objectives of establishing cost benchmarks and addressing the quality of the management of large projects were not met.
- (d) As regards the evaluation of the impact of the measures, the evaluation was based primarily on the use of aggregate macro-models which presents certain limitations. The specific macroeconomic model used (HERMIN) did not permit the necessary adjustments to take account of the specificities of the economies being evaluated (such as the existence of an important tourism sector). In addition, there were critical limitations in the estimation of certain key variables (such as was the choice of proxy data taken from studies of the US

economy to calculate external elasticities in lieu of actual data for the relevant European regions, which were not available).

As regards the Objective 3 evaluation, it was in part based on unreliable or incomplete data while there were significant mismatches between the data provided and the commentary thereon. Furthermore, there was a lack of convincing support for some of the hypotheses advanced thereby shedding doubts on the methodology employed to reach them.

117. Many of the shortcomings of the evaluations can be traced to the circumstances in which they were carried out, notably the limited information available and the tight time constraints set. It should have been possible to anticipate these constraints and the evaluation exercise should either have been given the necessary resources or the scale of the exercise should have been limited to what was more feasible. Furthermore, the Commission's oversight of the evaluation process needs to be improved.

118. The *ex post* evaluations have still provided some worthwhile insights, notably into the management systems underlying the implementation of SFs. However, the Commission encountered serious difficulties in delivering the desired in-depth analysis of the effectiveness, efficiency and impact of the funds, largely due to limitations in the approach adopted and restrictions in the availability of data. The latter constraint had not been adequately anticipated at the start of the exercise. These problems were compounded by certain weaknesses in the Commission's oversight of the process. Consequently, while some worthwhile and useful recommendations were made as input for future programming periods, the Commission could not benefit fully from the experiences of the 1994 to 1999 programming.

RECOMMENDATIONS

119. A reappraisal of the scope, procedures and approach used in *ex post* assessments needs to be done with some urgency, before the next batch of contracts for *ex post* evaluations are issued at the end of the current programming period 2000 to 2006.

120. To improve the evaluation process, better quality control procedures need to be introduced and effectively applied by the Commission so that the problems encountered in the evaluations under review do not recur in future *ex post* assessments. Such procedures should ensure:

- (a) that relevant and reliable data is collected regularly and is available at each stage of the evaluation process;
- (b) that terms of reference are adequate and provide for the application of appropriate methodologies by the contractors;
- (c) that adequate resources and time are allocated to the evaluation processes;

- (d) that there is appropriate monitoring and supervision by the Commission so as to ensure the overall quality of the process.

121. Secondly, particular attention needs to be reserved to the choice of appropriate techniques for assessing and measuring the economic impact. If a macroeconomic modelling approach is used to analyse the impact of the structural funds, it must be adapted to take due account of the characteristics of the economies under review. This consideration is particularly relevant in the situation where the evaluation will need to be extended to cover the 10 new Member States. In this light, the Court considers that there should be a re-appraisal of the appropriateness of an aggregate macroeconomic modelling approach (e.g. such as one based on the HERMIN model). The Commission should examine whether greater emphasis could be placed on models that rely on the use of micro-data derived from projects. It is also likely that a more in-depth evaluation of certain, specific themes could better lead to the identification of best practices and provide additional insight into particular issues such as for instance the impact of clustering, the significance of growth hubs in regional development vis-à-vis the overall objective of convergence.

122. Thirdly, greater emphasis needs to be placed on establishing linkages between the *ex post* assessments and thematic studies, and to ensure that results are consistent with each other. If results appear to be incompatible, then this aspect needs to be specifically tackled. In the evaluations under review, where linkages were developed by the evaluators, the quality of the assessment improved markedly.

123. Fourthly, a number of recommendations for future assessments more particularly for Objective 1 Regions may be made, namely:

- (a) There needs to be specific focus on the reasons why private sector contributions to the structural funds projects differ

markedly between the Member States. This may cast important light on the different drivers to economic development in the various Objective 1 regions.

- (b) Another worthwhile topic to explore is whether there should be a shift in the future from grant expenditure to financial engineering measures, such as seed and venture capital funds, loans and interest rate subsidies, which may prove to be more sustainable and efficient in the long term.
- (c) Greater attention needs to be given to the establishment of unit costs and benchmarks for various types of projects.
- (d) Additional attention also needs to be given to the project application and approval process as the evaluators pointed out.
- (e) There should be a greater focus on internal evaluations by the relevant ministries or regional authorities in future *ex post* assessments.

124. Finally, it is important to strengthen the capacity of the Commission to oversee the evaluation process, including the application of complex macroeconomic models. One possibility to consider, in this context, would be for the relevant DGs to seek a closer collaboration with research institutes and universities specialised in this area of assessment. Appropriate and timely supervision of contracted consultants is crucial in ensuring that objectives are met and that value for money is derived from such *ex post* assessments.

125. In conclusion, the *ex post* evaluation exercise must go beyond the compilation of distinct assessments, carried out at given intervals of time, but should be envisaged and managed as an integral part of an overall process. This is another aspect where a closer, and long term collaboration with research institutes or universities could be of significant benefit in providing a more effective framework within which to integrate the specific reports and recommendations generated at each stage of the evaluation process.

This report was adopted by the Court of Auditors in Luxembourg at its meeting of 20 July 2006.

For the Court of Auditors

Hubert WEBER

President

ANNEX I

Breakdown of aid under the Structural Funds — Available resources related to CSF 1994-1999

	(million euro)															
	Belgium	Denmark	Germany	Greece	Spain	France	Ireland	Italy	Luxembourg	The Netherlands	Portugal	United Kingdom	Austria	Finland	Sweden	Total
O1	748,50		13 985,30	14 333,90	26 965,80	2 245,40	5 762,30	15 236,20		153,80	14 333,90	2 419,70	171,50			96 356,30
O2	349,40	123,00	1 604,80		2 474,80	3 866,40		1 498,20	15,40	666,20		4 693,40	104,40	165,40	189,30	15 750,70
O3/O4	476,50	308,40	1 990,00		1 888,60	3 282,20		1 757,40	23,60	1 105,70		3 460,50	408,30	354,10	537,60	15 592,90
O5a Fisheries	25,10	143,20	76,30		122,20	194,40		137,60	1,10	47,70		90,80	2,10	23,80	41,40	905,70
O5b	78,90	55,30	1 257,30		680,40	2 293,30		923,30	6,10	153,70		837,20	424,90	201,00	142,70	7 054,10
O6														475,90	260,70	736,60
Total	1 678,40	629,90	18 913,70	14 333,90	32 131,80	11 881,70	5 762,30	19 552,70	46,20	2 127,10	14 333,90	11 501,60	1 111,20	1 220,20	1 171,70	136 396,30

Source: Special Report No 16/98 (OJ C 347, 16.11.1998).

Breakdown of aid under the Structural Funds —
Available resources related to Community Initiatives 1994 to 1999 ⁽¹⁾

	(million euro)
Interreg II	3 562,3
Leader II	1 777,2
ADAPT	1 646,4
PME	1 092,5
URBAN	894,5
PESCA	301,5
Rechar II	465,0
Resider II	610,9
RETEX	609,2
EMPLOI	1 858,5
Konver	734,9
REGIS II	615,0
PEACE	303,1
Total	14 471,0

⁽¹⁾ Names of programmes related to Structural Funds; further references can be found in the Commission budget, subsection B2.

Source: Special Report No 16/98 (OJ C 347, 16.11.1998).

ANNEX II

GDP and the employment rate

1. A multiple regression analysis was conducted to test the statement in the Synthesis Report that 'increased levels of GDP per capita have generally not been the result of increased employment rates resulting from job creation' ⁽¹⁾.

2. For strict comparability, the same data published in the Synthesis Report has been used in this analysis. In the case of the dependent variable the percentage change in GDP per capita for each Objective 1 region between 1993 and 2000 was used, while as main explanatory variables, the percentage change in the employment rate as well as that of the unemployment rate was used for the same regions, and over the two years 1993 and 1999 as published in the report. The results are given in *Table 1*.

Table 1 ⁽¹⁾**Regression of the change in GDP per capita on change in employment (CEMPL) and change in unemployment (CUNEMP)**

Regressor	Coefficient	Standard error	T-ratio	[Prob]
C1	1,7615	1,2361	1,4251	[0,161]
CEMPL	0,24563	0,14237	1,7252	[0,092]
CUNEMP	0,051927	0,027418	1,8939	[0,065]
BE	- 16,8803	6,2156	- 2,7158	[0,009]
IR	38,2350	6,5314	5,8540	[000]
DX	26,9660	2,9504	9,1398	[000]
R-squared	0,74449	R-bar-squared		0,71546
S.E. of Regression	6,1411	F-stat. F (5,44)	25,6414	[000]
Mean of dependent variable	6,3647	S.D. of dependent variable		11,5126
Residual sum of squares	1 659,4	Equation log-likelihood		- 158,5015
Akaike information criterion	- 164,5015	Schwarz Bayesian criterion		- 170,2376
DW-statistic	2,3210			

⁽¹⁾ 50 observations used for estimation. BE, IR and DX are dummy variables for Belgium, Ireland and Germany. C1 is the intercept.

3. From a preliminary assessment of the data, it emerged that three particular regions exhibited significantly different patterns of change in GDP per capita over the period as compared to the other regions, and which required specific treatment in the regression, as otherwise it would be difficult to draw general conclusions about underlying links. These were; Ireland, which saw its GDP per capita rise from 81 % to 115 % of the EU average over the period; the Hainaut region in Belgium, which saw its GDP per capita decline from 82 % to 71 %; and German Objective 1 regions which saw the GDP per capita rise from a range around 50 % to close to 70 % of the EU average by the end of the period (for example, Thuringen saw its GDP per capita rise from 52 % in 1993 to 70 % in 2000). To take account of such large relative changes in GDP per capita, specific statistical (or dummy) variables were included for this purpose which allowed the function to shift for these three broad regions.

4. The regression results shown in *Table 1* provide a number of relevant points worth noting. Overall, 74 % of the variation in the change in GDP per capita is accounted for by the 'explanatory' variables. Both main explanatory variables, that is the change in employment rates (CEMPL) and change in unemployment rates (CUNEMP) are statistically significant. In the case of CEMPL, the coefficient estimated is 0,24, suggesting that each one percentage point increase in the employment or participation rate is associated with an improvement in GDP per capita of Objective 1 regions by a quarter of a percentage point — thereby, if nothing else changes, closing the GDP gap by this amount.

⁽¹⁾ Synthesis Report, p. 62.

5. This result directly contradicts the conclusion reached in the synthesis report, and provides a useful indicator of the potential there is, from an analysis of the results of Objective 1 regions over this period, to close the GDP gap by improving participation rates — a most relevant result from the policy perspective.

6. A second feature of note in the results in *Table 1* relates to the impact of the changes in unemployment rates on changes in the GDP per capita. This impact (also discussed in the next section) is also statistically significant, but is rather small quantitatively, almost of an irrelevant value. Unemployment declined marginally overall in Objective 1 regions, and this may be the reason for this result.

7. To explore further the robustness of the main result in *Table 1* relating to the impact of changes in the employment rate and changes in GDP per capita, *Table 2* provides the regression results for all regions in the EU-15 and for which data was consistently available from Eurostat, and not just for Objective 1 regions. The results from 145 regions were derived from slightly different data than in *Table 1*. Rather than taking just the rate of change between the beginning and end of the relevant period, the average rate of change was computed for the whole period. This was judged to give a more representative summary of developments over a number of years, in this case over the five year period between 1996 and 2000. (Due to data limitations in the available statistics, an earlier starting point was not feasible for a number of regions.)

Table 2 ⁽¹⁾

Regression of average change in GDP per capita on average change in employment (ACEMPL) and average change in unemployment (ACUNEMP)

Regressor	Coefficient	Standard error	T-ratio	[Prob]
C	5,2110	0,12999	40,086	[000]
ACEMPL	0,24119	0,63540	3,7959	[000]
ACUNEMP	- 0,0019479	0,012260	- 0,1588	[0,874]
ID	4,7009	0,85873	5,4743	[000]
LD	5,4309	0,84670	6,4142	[000]
BED	- 0,64276	0,25885	- 2,4831	[0,014]
DED	- 1,3747	0,17517	- 7,8479	[000]
R-squared	0,62600	R-bar-squared	0,60974	
S.E. of regression	0,84193	F-stat. F (6,138)	38,4979 [000]	
Mean of dependent variable	5,2359	S.D. of dependent variable	1,3477	
Residual sum of squares	97,8204	Equation log-likelihood	- 177,2101	
Akaike information criterion	- 184,2101	Schwarz Bayesian criterion	- 194,6286	
DW-statistic	1,6353			

⁽¹⁾ 145 observations used for estimation. ID, LD, BED, DED are dummy variables for Ireland, Luxembourg, Belgium and Germany. C is the intercept.

8. The results indicate that the coefficient for the average change in employment is practically the same as in the previous case, that is 0,24, indicating that for all regions, a 1 % increase in the employment rate tends to be associated with a 0,24 % increase in GDP per capita. This result is also statistically significant, as is the whole regression equation. (An additional dummy variable for Luxembourg has been added, again because the increase in GDP per capita from 161 % in 1996 to 199 % in 2000, is markedly different from the average performance of the other regions).

9. The impact of the change in the unemployment rate variable is again minimal and in this case is statistically insignificant and negative. The implications from these results are that the impact of changes in the unemployment rate on changes in GDP per capita does not appear strong — possibly because the unemployed are already accounted for in the labour force and in the measure of GDP that is used, and any change from unemployed to employed has a less easy impact to capture in statistical analysis of this type. Certainly, more research in this area seems warranted.

ANNEX III

GDP growth and the structural funds

1. A multiple regression analysis was conducted to explore the link between the average annual change in GDP per capita for the Objective 1 area (the dependent variable) vis-à-vis the actual EU funds per capita expenditure in the Objective 1 regions of each member state. The same data and time period (1993 to 1999) is used as that found in the Synthesis Report.

2. It should be noted that modelling rates of change is a statistically demanding exercise, and often gives low levels of explained variation in the dependent variable. Nonetheless, the results in *Table 1* show that 53 % of the variation in the dependent variable is accounted for by the different allocations in EU Funds on a per capita basis. This is a statistically significant result, and suggests that regions which have had a higher actual EU Structural Fund expenditure on a per capita basis, also tended to have a higher average GDP growth per capita over the period.

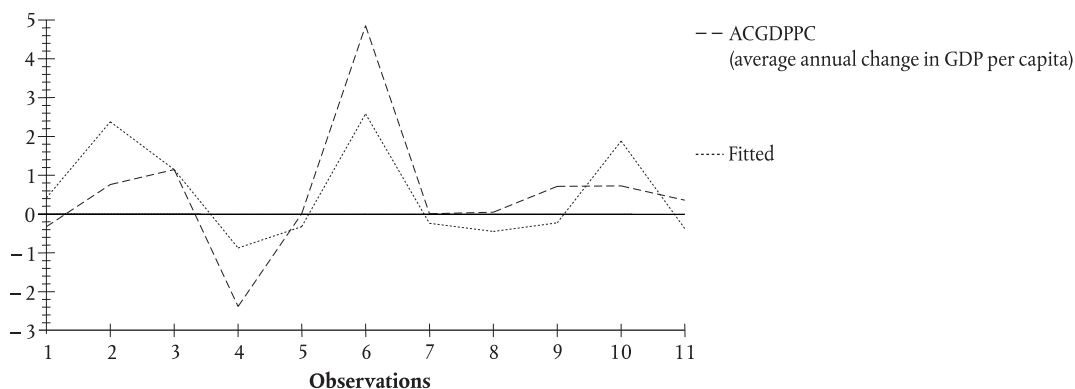
3. Although the sample is small, and this makes the statistical assessment difficult, each of the 11 observations is an amalgam of the information in many regions and thus represents a wider information base than the mere 11 observations would suggest. This result contrasts with that given in the report and suggests that further research in this area is warranted. The plot of the actual and fitted values from this regression is also shown in *Figure 1*.

Table 1

Regression of average change in GDP per capita on average change in EU structural funds per capita (EUSFPC)

Regressor	Coefficient	Standard error	T-ratio	[Prob]
C	- 2,0780	0,90194	- 2,3040	[0,047]
EUSFPC	2,8192	0,88559	3,1834	[0,011]
R-squared	0,52963	R-bar-squared		0,47737
S.E. of regression	1,2355	F-stat. F (1,9)	10,1340	[0,011]
Mean of dependent variable	0,53688	S.D. of dependent variable		1,7089
Residual sum of squares	13,7371	Equation log-likelihood		- 16,8304
Akaike information criterion	- 18,8304	Schwarz Bayesian criterion		- 19,2283
DW-statistic	1,8219			

Figure 1

Plot of actual and fitted values

ANNEX IV

HERMIN simulations and unemployment

1. The HERMIN model was applied on the basis of an assumed 'average' hypothesis (medium-medium) concerning supply effects or elasticities with respect to human capital and infrastructure. The robustness of these results were analysed by means of a sensitivity analysis on the basis of two further hypotheses:

- zero-zero: — minor neo-classical effects (via changes in relative prices),
- no supply effects (in terms of improvements in infrastructure and human capital) or small output and factor productivity elasticities with respect to physical infrastructure and human capital;
- high-high: — supply effects much more pronounced (in terms of physical infrastructure and human capital) and continue to apply following the CSF period.

2. As regards to unemployment, the HERMIN simulations for eastern Germany, Spain, Greece, Ireland and Portugal have led to changes in the unemployment rate which varies according to the three hypotheses. To assess better the appropriateness of the results, the Court converted the falls (or increases) in the unemployment rate into net jobs created (or lost). The results of converting falls (or increases) in unemployment into net jobs created (or lost) are given in *Table 1*.

Table 1

HERMIN simulations

	Net jobs created		
	Zero-zero hypothesis	Medium-medium hypothesis	High-high hypothesis
Eastern Germany	822 308	729 923	635 755
Spain	1 051 135	241 789	– 806 026
Greece	303 034	190 825	65 969
Ireland	86 679	64 621	36 335
Portugal	497 445	250 502	– 253 299

3. The validity of the results needs to be assessed in the light of the following observations:

- (a) In the case of the new German *Länder*, although the bottom-up data are still incomplete, it can be concluded that the figures produced using HERMIN for net jobs created appear, in all the hypotheses, substantially to exceed the data produced by monitoring directly the projects themselves (about 500 000 created jobs or maintained) ⁽¹⁾.
- (b) As regards Ireland, the comparison with other data show that the net figure of 64 621 that is attributed to the CSF appears somewhat understated compared with the other evidence. Monitoring figures show that the 'industry' OP led to about 212 000 gross jobs created and 90 000 net jobs created. Furthermore, the total number of jobs, across the whole economy, increased by 395 075 ⁽²⁾. The evaluators also note that it is important to look beyond the CSF in order to take account of policy elements, such as incentives for FDI and the policy of curbing wage inflation, that accompany the CSF. In other words, the model does not take account of these elements, which increase the impact attributed to the CSF.
- (c) The Spanish O1 CSF, which is the largest of the CSFs, by applying the average hypothesis, is far less successful than the new *Länder* O1 CSF in terms of job creation and outperforms the latter in this area only if it is supposed that demand effects predominate. However, under that hypothesis, the net figure for jobs created during 1994 to 1999 as a result of the O1 CSF would account for around 40 % of employment growth in Spain between 1995 and 2000, which appears rather optimistic.
- (d) Applying the average hypothesis and the hypothesis of dominant demand effects to Portugal gives net job creation estimates that are far higher than those produced by the national authorities ⁽³⁾. However, this has not been addressed.

⁽¹⁾ Evaluation report for Germany (O1), (pp. 42, 48, 70, 72, 85, 86).

⁽²⁾ Evaluation report for Ireland (pp. 61, 92).

⁽³⁾ Department for Forecasts and Planning (DPP).

- (e) The same is true for Greece, where the national authority's net job creation estimates substantially exceed the estimates obtained from the HERMIN model (under all hypotheses) concerning falls in unemployment. According to the national report, the CSF has contributed to an annual net increase in employment of 1,7 % of the labour force ⁽¹⁾ which translates as a net figure of more than 400 000 jobs created during 1994 to 1999.
 - (f) The figures for jobs lost under the high-high hypothesis for the Spanish and Portuguese CSFs do not appear credible.
-

⁽¹⁾ National report, p. 210.

THE COMMISSION'S REPLIES

SUMMARY

VI. The Commission is fully aware of the limitations of the study, due to a large extent to data constraints, and will take steps to improve the feasibility of future *ex post* evaluations.

VII. Although some difficulties with data were encountered, the evaluation used a number of types of information to reach its conclusions (member state mid-term and final evaluation reports, interviews, national workshops, and case studies).

VIII. The Commission considers that the terms of reference were adequate for the purpose, but the evaluation faced major challenges, especially of data availability. The Commission took into account both positive and negative aspects of the work carried out.

Nevertheless, the Commission accepts that a preparatory analysis would have helped to identify some of the information difficulties that the Court has described.

X. Although difficulties were encountered, the results of the evaluations have been considered during the mid-term review process as foreseen.

XI. The Commission considers that the future *ex post* evaluations should be more focused and take full account of data and resource constraints.

The Commission considers that macroeconomic models are a relevant tool for measuring the economic impact of large scale interventions. The HERMIN models developed for the various countries concerned took account as far as possible of the national specificities of the recipient economies. However, as for any evaluation tool, further improvements can be made to address limitations, in particular on micro data.

The Commission agrees that *ex post* assessments should take full account of existing thematic studies.

Future evaluations should cover fewer evaluation questions. This will make it possible to deepen the analysis in certain fields and will facilitate the oversight carried out by the Commission.

INTRODUCTION

1. The Court has made observations on the evaluation in general and *ex post* evaluation in particular on several previous occasions, notably in Special Report No 15/1998 on the assessment of Structural Fund intervention for the 1989 to 1993 and 1994 to 1999 periods and Special Report No 7/2003 on the implementation of assistance programming for the period 2000 to 2006 within the framework of the Structural Funds. In this earlier work the Court acknowledged the specific constraints of such

evaluations and came to an overall satisfactory assessment, while pointing out aspects needing improvement. The Commission welcomes the Court's constructive criticism on this topic.

4. For the future *ex post* evaluation the Commission will review the organisation of the exercise, with particular reference to data collection and the preparation of the synthesis report.

The ESF *ex post* evaluation used the evaluation reports carried out by Member States at both the mid term and 'final' stages of the programming period 1994 to 1999.

EX POST ASSESSMENT OF OBJECTIVE 1 BY DG REGIO

15. Text box 1. The *ex post* evaluation is not an official report of the Commission. Nevertheless, the Commission supports the overall conclusion of the report, which is borne out by evidence from the Commission's Third Report on Economic and Social Cohesion.

16. The Court acknowledges the particular challenges facing *ex post* evaluation, and in particular the lack of quantified indicator data. The Commission agrees that there is room for improvement in future *ex post* evaluations. To this end, the Commission intends to initiate a broader methodological debate with Member states and the academic world before the next round of *ex post* evaluations. A major issue that needs to be explored is whether to focus on a small number of relevant evaluation questions. The exercise will be less subject to data constraints.

17. Although the convergence rate differs between regions, growth in Objective 1 regions is driven not only by employment increases but also significantly by increases in productivity. Other empirical studies suggest that productivity is the key factor determining the growth rate that is sustainable in the long run (unless there is a permanent inflow of migrants).

18. The Commission welcomes the regression analysis undertaken by the Court. However, it is based on only one of the explanatory variables. To avoid biased results, a control analysis for the other explanatory variables would be necessary.

19. Other studies undertaken by the Commission, for example for the Third Cohesion Report, have also led to different conclusions from the Synthesis Report. The picture is in fact mixed, with increases in employment appearing to be the main factor in growth in some Member States, and productivity growth in other Member States (see Third Cohesion Report, p. 3).

20. The Commission agrees that the data available on participation rates, gross domestic product (GDP) and productivity could have been better exploited by the evaluator. However, other studies have also explored these relationships.

21. Planned and actual expenditure data cannot be broken down by region, as a significant share of the assistance is delivered through national programmes such as transport, research, national aid schemes or human resource programmes, and Member States do not keep statistics on how it is allocated by region. Therefore, the regression could necessarily only cover part of EU expenditure. This was an inherent constraint on this part of the evaluation.

23. In its Third Cohesion Report (p. 147) the Commission showed that there was some relationship between the amount of structural aid provided and real growth of GDP in Objective 1 regions.

24. As already pointed out, the evaluation was faced with severe data constraints, particularly regarding national data on both public and private co-financed expenditure. The data presented were mostly based on estimation procedures, which gave rise in some cases to errors or discrepancies.

26. The participation of the private sector depends on the general economic conditions of the country as a whole and the region concerned. The leverage effect of EU Structural Funds varies greatly from one country to another.

27. Examination of this issue was hampered by the uncertainty surrounding figures for the private sector.

28. Market failure is a crucial issue in the context of *ex ante* evaluation. Although it is mentioned in the methodology guide drawn up by the evaluator, a formal and detailed examination of appropriateness was not required by the terms of reference.

30. The evaluators had to base their judgements largely on information and data available in the Member States. If data, for example on deadweight, was not available, the terms of reference did not require the evaluators to collect such data themselves, given the limited budget available for the exercise.

Concerning the concept of Community value added, the Commission considers that the content of the concept depends on the context in the different Member States. For example, the programming approach can be new in one Member State but not in another. A qualitative assessment seems the most appropriate in these circumstances.

31. The limitations are largely due to the data and budget constraints. While a consideration of the results of other thematic studies may well have given added weight to the conclusions, the *ex post* evaluation was necessarily broader in scope.

32. The two evaluations have a different scope and coverage. For example, the thematic evaluation of transport is also based on the Cohesion Fund and covers all objectives.

33. Estimating the impact of Structural Fund interventions on employment is notoriously difficult and depending on the methodology can produce different results. These should therefore be interpreted with caution. The Commission is doing further methodological work on the estimation of employment effects to increase the consistency of results.

34. The methods used for estimating the employment effects in the Synthesis Report and the thematic evaluation report are not comparable. The HERMIN model is based on more robust assumptions and also captures supply-side effects associated with the investment in physical and human capital, while the study on transport estimates the direct and indirect short term effects.

35.-36. As indicated above, estimating the impact of Structural Fund interventions on employment is notoriously difficult and depending on the methodology can come to different results. The thematic study on SMEs carried out a more in-depth analysis on SME interventions and also covered other interventions outside Objective 1 areas.

37. Monitoring data of the Member States was a necessary input for the study. Given the time and budget constraints, it would have been impossible to make up for the gaps and deficiencies in the data.

38. Given the broad scope of the evaluation, it was impossible to address all issues in depth by undertaking ad hoc surveys or other field work.

39. The Commission agrees that evaluations should use existing studies and literature as a starting point for the analysis. However, the thematic studies had been undertaken in another context, namely as an input for preparing 2000 to 2006 programmes, whereas the *ex post* evaluation came later.

40. The use of formalised quantitative methods requires the existence of robust and sufficient data from the Member States. As the Court points out, such data was often missing for the 1994 to 1999 programming period.

41. The objective set by the terms of reference (ToR) for this area was fairly limited in scope.

42. Despite the difficulties encountered in obtaining data due to the fact that different types of building work were interconnected, the evaluators attempted to establish comparable unit costs for some fields such as road construction, environmental projects and jobs created in industrial projects.

43. The Commission agrees that a more detailed analysis is desirable. The subsequent *ex post* evaluation on the Cohesion Fund delivered further insights on this subject. In 1998 the Commission prepared a guide entitled 'Understanding and monitoring the cost-determining factors of infrastructure projects — A user's guide'.

44. The wording of the ToR made it clear that these tasks were especially difficult and could only yield partial conclusions. An assessment of public-private partnerships and of tendering procedures was not required by the ToR. The issue of public-private partnerships has been addressed in the *ex post* evaluation of the Cohesion Fund.

46. The evaluator excluded private cofinancing in the macroeconomic modelling, as the model sought to quantify the overall impact of EU and public cofinanced expenditure on the economy including private investment. Private sector investment is regarded as a result and not as a separate policy input. The Commission refers to points 24 to 27 of its replies. The Commission shares the concerns of the Court on the difficulty of specifying externality parameters. The choice made was based on international studies from which a range of elasticity values could be derived. At the time the study was conducted, there was no real alternative for conducting the impact analysis.

48. Some of the Court's observations in points 48 and 49 were made in its Special Report No 7/2003 on the implementation of assistance programming for the period 2000 to 2006 within the framework of the Structural Funds. The Commission refers to the replies given in that report.

The alternative suggested by the Court was not feasible. The requirements for collecting micro-data were not standardised to the extent that would have been necessary for this type of analysis. The Commission agrees that future macroeconomic evaluations should rely more on microeconomic data analysis.

49.

(a) These were the assumptions of the HERMIN model at the time of the evaluation. However, further development of the model should pay greater attention to the importance of the service sector in Objective 1 economies.

(b) The issue of structural instability of the model will be examined carefully in future macroeconomic work.

(c) The Commission acknowledges that the modelling exercise would have been more accurate if data sets on actual expenditure had been available. The commitment data used were the only data available across all regions at the time of the evaluation.

(e) The elasticity values associated with physical and human capital investment which are used in the HERMIN model are based on a broad range of academic studies. The model

considers different scenarios based on high, medium and low values. This sensitivity test results in a range of impact values.

All macroeconomic models face data problems. The modelling exercise relied on the best data available. The Commission agrees that additional work is needed to further improve the elasticities data used by the model.

(f) In the HERMIN model, the counterfactual scenario is defined as the situation with no CSF in order to be able to estimate the deviation relative to the baseline. However, other additionality scenarios can be envisaged, e.g. only structural and cohesion funds (without national co-financing).

50. In the Commission's opinion, despite possible shortcomings, HERMIN-type models are a valuable instrument for understanding and measuring the effects of cohesion policy. The use of the model by a growing number of Member States and the absence of convincing alternative approaches confirm this view. The HERMIN model can be improved and the Commission is working on its development for the programming period 2007 to 2013.

51. The Commission considers that it is difficult to establish any linear correspondence between a fall or increase in unemployment and the number of jobs created or lost. Labour market mechanisms are more complex and should be examined carefully before drawing firm conclusions.

52. Again, the evaluation was dependent on the quality of data and setting of objectives by the Member States. The absence of such data made the evaluation in many cases difficult or impossible.

56. Partially conflicting or differing recommendations in different evaluations with different questions and methods are a normal and productive phenomenon in economic and sociological research.

57. The recommendation in the *ex post* evaluation to pay attention to an appropriate, workable size of monitoring committees is not inconsistent with the recommendation to ensure inclusiveness in their composition made in the thematic study.

59. The *ex post* evaluation was based here only on partial studies. The Commission agrees that the findings concerning the management systems were interesting enough to deserve more attention. A specific evaluation of this issue was undertaken by the Commission services in 2003 (*The efficiency of the implementation methods for Structural Funds*, December 2003).

60. Like the other thematic evaluation studies, that of the partnership principle was designed to inform the programming of the 2000 to 2006 period. The programming had been completed by the time of the *ex post* evaluation. Therefore, greater discussion of this subject in the *ex post* evaluation, while it may have been desirable, was not essential.

61. The tasks and budget of the evaluation were necessarily limited.

The issues raised by the Court were the subject of the study referred to in the reply to point 59. Project selection procedures are also one of the aspects of management and control systems that are checked in audits, as they are an important condition of sound financial management.

The Commission tends to the view that future *ex post* evaluations should cover less questions.

62. The Commission agrees that Member States' management and control systems for the Structural Funds during the 1994 to 1999 period contained weaknesses. However, they improved towards the end of the period with the introduction of the control regulation 2064/97, and the improvement has continued in the 2000 to 2006 period.

63. Member States and Commission agree that the monitoring of Structural Funds programmes was a weak point in the 1994 to 1999 programming period. In the 2000 to 2006 period, substantial efforts were undertaken to address this issue. For the first time systematic evaluations (including *ex ante* evaluations and mid-term evaluations) were carried out for all Structural Funds programmes. This led to a substantial improvement of the indicator and monitoring system for the Structural Funds.

64. The financial data produced by the systems were broadly reliable. This was not, however, the case for monitoring data, although the availability of such data improved in the later stages of the programme period.

65. 'Internal evaluations' were not legally required by the regulations but were performed in response to specific needs felt by the Member States. This was a positive by-product of the *ex post* evaluation, which is now being widely imitated.

66. For the programming period 2000 to 2006 the regulation established for the first time detailed rules and requirements. Based on that, the Commission elaborated detailed working papers on the elements of monitoring and evaluation. The *ex ante* evaluations, carried out systematically for all programmes and objectives, were a major opportunity for the Member States to invest in the establishment of indicators and quantified objectives. Most Member states introduced IT solutions to support regular and systematic reporting based on quantified information. The mid-term evaluation was an opportunity to revise and to improve the indicators and their use.

67.

(a) The Terms of Reference stated that 'data available in the Member States may not be complete, particularly if the Member State does not plan to undertake *ex post* evaluations...'.

(b) The extent of further data collection was constrained by the time and resources available.

(c) There are methodological difficulties in assessing net effects through microeconomic methods. However, further work should be undertaken in this area.

(d) The Commission agrees that this is an important issue in assessing effectiveness. However, it would have required specific further studies, which are expensive. The Commission, however, encourages Member States to undertake such work.

(e) The potential links between strategy, effectiveness and impacts are analysed in the *ex ante* evaluations.

(f) and (g) In the Commission's view, it is better in terms of reference to set only the evaluation questions, leaving a methodological proposal to the bidders.

68. A great deal of research on convergence has been done. The results of recent such work are summarised in the Commission's Cohesion Reports.

69. This could have been another interesting approach, but was not necessary. In the Commission's view, future evaluation must set clear priorities, limiting the scope of questions.

70. The Commission considers that the ToRs and the methodological guide were adequate for the purpose.

72. Learning from the *ex post* evaluation, the Commission currently uses expert advice in methodologically demanding evaluations at an earlier stage.

73. The *ex post* evaluation 1994 to 1999 faced specific challenges, especially of data availability.

75. The Commission applied in this evaluation the usual management techniques such as a formalised interim reporting system, oversight by a steering committee, regular working meetings and the involvement of experts. A report of acceptable quality was the outcome of this process.

The complexity of the *ex post* evaluation was significant. The Commission will use the experience gained to address this issue better in future.

The Commission agrees with the view of the Court that future evaluations should cover fewer evaluation questions. This will make it possible to deepen the analysis in certain fields and will facilitate the oversight carried out by the Commission.

76. The evaluation, both in the Executive Summary and Chapter 9 Recommendations, goes far beyond management recommendations. The recommendations are aimed at Structural Fund interventions in all Objective 1 regions and are necessarily of a general nature. The role of research and development is stressed, as are the need to pay more attention to differences within large Objective 1 regions, the need to coordinate better Structural Funds interventions across funds in rural areas, etc.

77. The Commission agrees with the evaluator that Monitoring Committees should devote more attention to strategic issues. This must not detract from the other tasks Committees have to perform under the regulations.

80. The recommendation specifically referred to by the Court has been taken up by the Commission for the 2007 to 2013 programming period.

EX POST ASSESSMENT OF ESF OPERATIONS (UNDER OBJECTIVES 1 AND 3) BY DG EMPL

81. A coordination meeting between DG REGIO, DG EMPL and the evaluators for the ESF *ex post* evaluation took place in October 2002. At this meeting it was agreed that the ESF *ex post* evaluators would have access to material gathered by the ERDF *ex post* evaluators. This later took place. The two managers of the *ex post* evaluations were members of each other's steering groups and received documentation circulated to other steering group members. The ESF evaluators used the ERDF evaluation material available.

Text box 2. A variety of sources of information was used by the ESF *ex post* evaluators.

82. The synthesis report alone contains a great deal of information concerning the use of ESF resources (nature, quantity and impact) across different target groups and by types of measure across the Member States (see in particular Chapter 2 of the Synthesis Report).

84. The number of beneficiaries estimated in the table mentioned is based on the information provided by the Member States' monitoring systems. The difficulties encountered in the accuracy of some data did not have a significant influence on the useful observations issued.

The beneficiary totals only play a role in relation to (i) accountability and (ii) the limited input/output relationships analysis.

The *ex post* evaluation did not rely on single sources of information in order to reach conclusions and make recommendations.

85. The Commission recognises that there were problems in estimating the number of ESF beneficiaries. Nevertheless, it considers that the table contains useful data for policy makers.

While the *ex post* evaluator has already, quite correctly, cited the two sources of information which have been used to develop the table, a fuller explanation of how the figures were derived would nonetheless have been useful.

86. The comments in the text refer to more detailed information than that presented in the accompanying table. The table presents averages while the text gives a breakdown by subcategory.

The Commission holds the view that the fact that a wide range of values is not provided does not detract from the value of the analysis.

87. The aim of Chapter 3 is to provide contextual information for ESF actions and it is in fact Chapter 2 that provides the bulk of numerical information concerning the ESF.

Text box 3. This limited number of textual mismatches does not invalidate the accuracy and usefulness of the data in the table.

88. The table clearly shows overall support for the hypotheses identified.

Examination of Table 3 shows that:

- (a) the highest number of cases which contradict any hypothesis is one out of 30;
- (b) at least two thirds of the scores are in the categories 'strong supporting evidence' or 'some supporting evidence' (20/30, 21/30, 22/30 and 26/30).

89. It is also a legitimate aspect of evaluation work to identify deficiencies in data.

The terms of reference clearly state that the evaluation should be primarily desk-based using secondary data (data and information available in the mid term and final evaluations and in the managing authorities' reports and monitoring systems) and that the evaluators should not engage in extensive primary information gathering. Significant additional field work and survey samples would have required extra resources and would have engendered significant additional costs especially given the need to collect information concerning the entire programming period. The forthcoming ESF *ex post* evaluation for 2000 to 2006 will be preceded by a preparatory analysis.

90. The synthesis report contains several references to the significance of 'combined measures' in terms of effectiveness.

92. Firstly, the evaluators choose their words with care and emphasise that the conclusions they draw are tentative 'The combined allocation of national expenditure and ESF resources *does not appear to have led* to a convergence ...'. Secondly, the figures (tables in Chapter 3 and Annex 2) do show that there is not a convergence of expenditure patterns between countries in combined allocation of national expenditures. Thirdly, as mentioned previously in reply to point 87 (including Text box 3) the data presented in the Chapter 3 tables mentioned by the Court are accurate and support the qualified observation provided by the evaluators. This observation is also based on the analysis of labour market trends presented in Annex 2 of the Synthesis Report (not only Chapter 3). The Commission has already indicated that the examples given concerned a limited number of textual mismatches which do not undermine the figures.

93. As in the case of Table 3 on individuals mentioned in point 88, the data clearly shows overall support for the hypotheses identified.

95. The Commission considers that its quality assessment was accurate and balanced, taking into account both positive and negative aspects of the work carried out (rather than only concentrating on a limited number of negative aspects).

Taking into account the above, and since the results and conclusions were based on a variety of sources of information the Commission stands by its quality assessment.

97.

(a) It is not surprising that some evaluation results are of a qualitative nature as the evaluation makes use of secondary sources which have been summarised and primary sources which are of a qualitative nature (interviews with stakeholders and national workshops on the evaluation findings). In some cases the description of results can only be of a qualitative nature. Estimation of deadweight was not a compulsory subject of analysis for Member States during the programming period 1994 to 1999, nor was it part of the evaluation terms of reference. The logical consequence of this is that the *ex post* evaluation itself could only report on deadweight estimates that were available to it.

(b) The UK report provides a substantial amount of evidence on soft outcomes.

Other national reports also make references to the concept.

(c) The synthesis report, and the national reports on which it is based contain numerous observations about the content of ESF action (financial, by target group and by types of measure) and performance, comments derived from national evaluation reports and comments derived from interviews with stakeholders).

Information on performance is also available from the case studies.

(d) Sustainable development was not yet a political priority for the Structural Funds at the time when the programmes for the 1994 to 1999 period were being established and therefore it was not identified as a theme in the terms of reference.

Socio-economic cohesion does not appear in the terms of reference: it was dealt with in the Cohesion Report.

(e) Each of the national reports examined by the Court contains substantial material on the long-term unemployed and young people. The synthesis report also contains information on the above two 'target groups'.

(f) As has already been stated, the *ex post* evaluators were in some respects dependent on information provided by Member States. The fact that some programmes lacked 'specific initial objectives' was a further constraining factor for the evaluators.

(g) The ESF and the ERDF *ex post* evaluators collaborated and the ESF *ex post* evaluators used reports provided to them by the Objective 1 evaluators. Productivity increases by ESF beneficiaries in terms of the benefits to the firms employing them were not a required subject for Member States in their evaluation work during the programming period 1994 to 1999, nor was it part of the terms of reference. As such it is not surprising that this subject occurs only occasionally.

98. The synthesis report identified numerous factors impacting on the development of Active Labour Market Policies (ALMPs), such as the level of ESF resources, the Structural Funds planning processes and organisational arrangements for the ESF. It also described rather concretely the influence of the ESF in the context of the European Employment Strategy (EES).

As already pointed out, net impact analysis was not a compulsory subject for Member States during 1994 to 1999. Therefore, it is normal that the subject does not appear in all Member States' evaluation reports. Since the evaluators were requested not to carry out large scale primary data gathering, the *ex post* evaluators cannot be criticised for not having provided this material.

100. The EES stronger strategic framework provides targets against which progress can be judged. Improved institutional structures have been created to measure progress. Moreover, the EES promotes stronger links between EU programme funding mechanisms and other policy instruments, regulations and national programmes.

101. The Commission believes that each national report has dealt with the aspects of Community added value identified in the terms of reference (financial and leverage effects, policy and institutional effects, socio cultural effects, and variation across objectives). It recognises, however, that these aspects were sometimes dealt with in other sections of the national reports.

In addition, the national *ex post* evaluation reports (and the synthesis *ex post* evaluation report) integrate these various aspects into an overall analysis of Community added value.

102. The national and synthesis reports contain a great deal of information concerning the financial resources that have been devoted to ESF action.

In the case of both Spain and Italy, the evaluators arrived at positive conclusions overall.

The *ex post* evaluator for Spain reports that: 'The general picture that emerges is that the ESF funds have been adequately and efficiently used by Spanish authorities in the 1994 to 1999 period. Most of the programmes have been fully executed in the way that had been planned' (p. 74).

In the case of Italy, the evaluator concludes that: 'Some of these costs are necessary to improve performance and transparency in the implementation process; others could probably be reduced by simplification of procedure'. (The issue of simplification of procedures was taken up in the next programming period 2000 to 2006).

104. The Commission believes that the report does indeed provide useful material for conducting future *ex post* evaluations.

The report contains rather comprehensive reviews of approaches used to evaluate ESF type actions, both in relation to actions targeted at individuals and actions targeted at systems.

With regard to approaches applied in assessing net impact on individuals, the report contains numerous references to techniques used to estimate net impact on individuals (see especially p. 14, although additional references can be found on p. 10, 12, 21, etc.).

105. As regards the recommendations on policy analysis, the methods report clearly shows that the evaluators have identified a number of concrete recommendations linked to the area of policy analysis, for instance close monitoring of the relationships between ESF interventions and the evolving national and other regional level interventions.

Furthermore, the methods report discusses the 'logic models' for policy analysis at some length and develops evaluation questions relating to the justification for ESF action.

The *ex post* evaluators were not expected to propose an approach for the selection of appropriate indicators at different strategic levels because such indicators had already been set for the entire

programming period in guidance documents issued by the Commission in 2000.

106. The Commission considers that the evaluation tasks are adequately specified in the TOR. Conclusions in national *ex post* evaluation reports are based on diverse sources (interviews, case studies, etc.). The synthesis report provides an overall analysis of the results contained in the national reports. The contractor has provided an explanation of how these sources of information were used to generate conclusions in the proposal submitted to the Commission.

Describing all aspects of ESF action and fixing all aspects of the evaluation *a priori* would lead to unmanageable and inflexible terms of reference.

With regard to 'concrete action' to evaluate the impact on individuals, when the indicators are not concretely established or when specific statistical evaluations have not been carried out recently, as already stated, the entire evaluation approach followed by the *ex post* evaluator was designed to draw conclusions from a variety of sources of information (Member State evaluation reports, national workshops, face-to-face interviews with programme stakeholders, case studies, etc.).

The extent of the mismatch between quantitative objectives and available data was not known at the time of the drafting of the terms of reference, therefore it seems incoherent to expect the terms of reference to have been drafted differently.

The main contractor developed, in cooperation with national contractors, operational guidelines for the work to be conducted at national level.

107. The terms of reference did not require the evaluators to assess the 'appropriateness of the reprogramming choices', but only to describe them. In each of the cases cited the *ex post* evaluator has described reprogramming changes based on the available information.

108. The possibility of a preparatory analysis is being considered for the next exercise.

109. The Commission considers that the data difficulties faced by national evaluators did not detract from the overall quality of the evaluation. The evaluation design foresaw the use of a variety of sources of information. In addition to reviewing the content of Member State evaluation reports, the national *ex post* evaluation reports include material from closure reports, case studies, national workshops and face-to-face interviews.

In terms of delays, the steering group was not in a position to speed up the closure of mainstream ESF programmes, which are dependent on implementation by the managing authorities.

At the start of the evaluation, the *ex post* evaluator was provided with all evaluation reports produced by the Member States during 1994 to 1999. Furthermore, the *ex post* evaluator was provided with letters of introduction in order to facilitate contacts with national ESF administrations.

111. In the future the Commission will take all necessary steps for the successful preparation and close monitoring and supervision of the exercise.

CONCLUSIONS

115.

- (b) The extent of further data collection was constrained by the time and resources available.

Nevertheless, a substantial amount of additional qualitative data was collected, e.g. from interviews and workshops. National closure reports and the evaluation reports conducted by Member States during the 1994 to 1999 programming period were also reviewed.

- (c) A variety of sources of information was used in order to arrive at balanced, well considered conclusions.

In some cases the description of results can only be of a qualitative nature. But this was not solely the case: the evaluation results also contain significant quantitative information.

- (d) The conclusions drawn were based on the evaluators' use of secondary sources (which they summarised) and primary sources of a qualitative nature (interviews with stakeholders and national workshops on the evaluation findings).

See also reply to points 88, 92 and 96.

- (e) An evaluation of deadweight was not foreseen in the ToR.

As a general remark, an evaluation exercise cannot cover all the topics and has to focus on priority objectives and areas taking into account regulatory obligations, time and resource constraints and the intended use of the results. The experience of this exercise and lessons learned will be used for the 2000 to 2006 *ex post* evaluation.

116.

- (a) Examination of this issue was hampered by the uncertainty surrounding figures for the private sector.
- (b) The analysis was constrained by the availability of data and lack of comparability resulting from differences in methods and scope.

- (c) Some attempt was made to assess efficiency on the basis of a limited number of projects.

- (d) The HERMIN models did take account of the national specificities of the economies, however further work will be undertaken to address the limitations.

The Commission considers that the limited number of drafting mismatches in Chapter 3 of the report do not call into question the accuracy, validity and usefulness of the data to support the conclusions drawn. The data contained in the tables commented upon by the Court in points 88 and 92 show that there is significant overall supporting evidence for the hypotheses identified.

The difficulties encountered regarding the exhaustiveness of some data did not have a significant influence on the useful observations issued.

117. The Commission considers that the evaluation relied on a robust design using a variety of information sources and that its conclusions are sound.

In the future, the Commission will take all necessary steps for the successful preparation and close monitoring and supervision of the exercise.

See reply to point 111 above.

118. In the future the Commission will take all necessary steps for the successful preparation and close monitoring and supervision of the exercise.

RECOMMENDATIONS

119. For future *ex post* evaluations preparatory analyses or feasibility studies will be carried out to assess the availability of data and what outputs can reasonably be delivered within the budget and timeframe.

120. The Commission agrees that the quality control procedures applied in the 1994 to 1999 period *ex post* evaluations should be further improved.

The Commission recognises that for the forthcoming Structural Funds *ex post* evaluation exercises a preparatory analysis could play a useful role in facilitating the work of contractors during the evaluations proper.

- (a) The basic data used in the *ex post* evaluation are collected and provided by the Member States through the monitoring system. The availability of reliable data may be a constraint on the evaluation, but in the 2000 to 2006 period there has been progress in this area.

- (b) Terms of reference must enable the Commission to assess the ability of the candidate to conduct the evaluation and whether the particular combination of methods and techniques proposed can deliver the required outputs.

The Commission shares the Court's opinion that terms of reference play an important role in achieving good quality evaluations results.

- (c) Given budget and time constraints, future *ex post* evaluations will be more focused on key evaluation questions and will be prepared and launched earlier than the 1994 to 1999 exercise.

Based on the experience of 1994 to 1999, monitoring and supervision arrangements will be strengthened and the process launched earlier.

121. The Commission considers that macroeconomic models are relevant for estimating the economic impact of large-scale interventions. The HERMIN model was designed specifically for such purpose in the early 1990s. However, as for any economic model, further improvements will need to be made, in particular regarding the micro-data for the estimation of the economic returns of physical and human capital investments. The Commission considers that the future *ex post* evaluation should be more focused.

122. The Commission agrees that *ex post* assessments should take full account of existing thematic studies. When using them, that *ex post* assessments and thematic studies serve different purposes and therefore may lead to different estimates.

123.

- (a) The Commission agrees that private investment is a key driver for economic development. However, experience shows that private sector leverage differs across countries depending on specific economic conditions.
- (b) This approach cannot be universally applied since economic conditions in Member States and regions differ greatly.
- (c) The Commission is aware that further systematic work is needed in this extremely complex area.
- (d) The Commission is aware that further work is also needed in this area.
- (e) *Ex post* evaluation reports commissioned by Member States were one of the primary sources of information in the 1994 to 1999 evaluations. Although the Commission is primarily responsible for *ex post* evaluation, cooperation with Member States should be strengthened in the future and internal evaluations could play a greater role.

124. The Commission is taking steps to improve its oversight of the process.

125. The Commission agrees that the SF *ex post* evaluations should make full use of knowledge gained during an entire programming period and for this reason accepts the idea that a preparatory analysis could contribute to facilitating the task of contractors and improving the quality of forthcoming SF *ex post* evaluations.

The Commission already collaborates extensively with research institutes and universities and is willing to envisage further assignments for the purposes referred to by the Court.