This Annex consists in:
(i) A detailed description of methodological faults found in earlier research studies. (See Chapter 1)

Annex 1. The Determinants of Export Performance in the UK Knitted Apparel Industry: Subsample Of Empirical Studies: Methodology Assessment:

1. Introduction: Chapter 1 outlined, inter alia, main shortcomings in the literature identified by Zou et al, 1998 and other scholars in the field. It also highlighted the fact that there had been a very limited assessment by these and other reviewers of the methodologies used to investigate the determinants of export performance. As the amount of trust that can be reposed in a research study depends centrally on the soundness of its research methodology, and given that specific flaws in this area had already come to light, it was decided to take a closer look at this aspect. Accordingly a 20 per cent random sample was selected from the 50 studies reviewed by Zou et al and this was subsequently supplemented by a further 8 judgemental selections from the literature. The range of methodological flaws identified in this exercise is believed to cast considerable doubt on the reliability of research findings on export performance. Our findings are described below; they are also summarised in Table 1.10. But, as a prelude to this process, it was also felt useful to set out, first, some commonsense criteria for such an assessment and, second, to highlight the general importance of minimising the magnitude of avoidable variability within and between survey research cases.

2. Some commonsense criteria for the evaluation of research study methodologies
For soundness in empirical academic studies it seems clear that all such projects should seek to ensure at least that they:
* comply fully with accepted academic standards and best practice in respect of research methodology;
* minimise controllable variability within the empirical data collected to meet the research objectives;
* minimise confusion and enhance comparability by   using standard definitions of the study eg for industry sectors and firm sizes.
* describe the methodology and conduct of the research in a manner detailed enough for the assessment of the adequacy of both; and,
* as with the experimental method, the   information provided should ideally be of a comprehensiveness sufficient to permit readers to replicate the study if desired.

3. Sample survey of research methodology: With this background, we investigated the methodologies   used in a sample of 18 export performance studies, of which ten were a simple random sample drawn from the list of fifty reviewed by Zou et al (Items 1 to 10: Table 1. 10) and another four (Nos 11 to 14) were post-1998 performance studies judgementally selected from the literature search. In 2007 a further four (Items 15 to 18) were selected in the same manner, from the same period, and subjected to systematic review. Zou et al's criticisms of the literature (fragmentation etc.) have of course already been mentioned (Chapter1). The chief shortcomings we identified in this group of studies fall into 12 categories. They are discussed individually below.

4. The problem of inter-sector variability: After outlining the benefits of minimising variability, this section goes on to argue that , if it is not dealt with, unquantified inter-sectoral variability can pose a threat to validity. Turning to the first, there is often much avoidable variability among survey population elements whose reduction can assist, in particular, with the making of comparisons between and generalising across empirical studies. Scholars have rightly stressed the importance of minimising it. Zou welcomed the contribution made thereto by the widespread use of mail surveys *(Zou, S. et al, 1998, p340).* Mejia-Gomez noted the benefits of the common trading environment shared by the firms in his US regional sample *(Mejia-Gomez, L. R. 1988, page 497).* Katsikeas et al set their study in the Greek food industry partly to maintain as low as possible the hetereogeneity which ' *...could significantly diminish the usefulness of study findings...' (Katsikeas, C.S., et al, 1996, p16).* Chetty's study was `*…confined to one manufacturing industry to avoid any cross-industry variances.' (Chetty, S. et al, 1999, p 131)* And, finally, Lim took a number of steps to enhance their completed sample's homogeneity *(Lim, J.S., 1996, p69).* Therefore it is clear that scholars recognise the key value of homogeneity and the importance of comparing like with like.

4.1. But they do not appear to have considered in any systematic way the effects of cross-sectoral sampling hetereogeneity. We suggest below that, at least in the case of samples drawn to test hypotheses about population parameters, the use of unadjusted cross-sectoral samples may undermine the validity of the related research.

4.2. This might be less of a potential problem if they were few. But the preponderance of multi-sectoral studies in the export performance field is illustrated, for example, by their weighting of 88 per cent in Zou et al's review sample *(Zou, 1998, op cit, p345 et seq);* and 66 per cent in the Leonidou study, Although we have not seen any scholarly discussion, it seems likely that this multi-sector sampling is usually done with the aim of increasing the generalisability of survey results, as Leonidou suggests *(Leonidou, L.C., et al, 1998, p6).*

4. 3. But it is contended here that unless the data from cross-sectoral studies are adjusted or weighted, in some suitable manner, to compensate for the variability between sectors (See eg Das, No 5), the results emerging may not be reliable. The reason offered is that such raw samples would comprise `apples and pears' sampling unit subgroups that differ too much to count as one sample.

4. 4. Do researchers, then, weight their samples to reflect the differences that exist between the subsamples they have drawn from different sectors?. That does not appear to have been the case with the studies comprising our subsample of 18. While some were sliced into employee size-subgroups, they were never split similarly by sector. None of them included explicit weighting nor was any reference made to a possible need for this. We have not been able to find any related research.

4. 5. It does seem clear, however, that there is at least implicit recognition, at both macro and micro level, that there is a problem here. For instance, Zou's EXPERF Scale *(Zou,S.,et al, 1998b, p40)* attempts to improve the comparability of studies across national borders*. And Aaby et al's stress on the need to control for each export market's stage of development is also about adjusting, from a different angle, for cross-border variability; while their reference to Reid's 1986 research more centrally points in the same direction. Also at the micro level, Zou et al's insistence on the need, in survey samples, to control for large diversified companies is surely likewise a recognition of inter-sector variability. Indeed, such appears to be almost explicit in their reference to studies that are `difficult if not impossible to compare'.

(Zou, S. et al, 1998, pp345-350): (Aaby, N-K. et al, 1989, p22,18). But it seems that the arguments used or implied in specific cases may not have been applied consistently right across the `coalface'. (* And it would seem to require rigorous argumentation to show that this level differs in kind and not merely in degree from the inter-sectoral).

4. 6. If our argument is sound, it would appear to have three specific implications for research, at least in this field. First is that unless suitable weighting of the component subsamples is conducted, the hoped-for gain in generalisability from multi-sector sampling may be largely or entirely offset by the unadjusted inter-sector variability- undermining, too, external validity. Second, cross-sectoral subgroups of sampling units may be too differentiated, one from another, to have a plausible claim to be the single uniform sample required to satisfy the minimum size requirement of the selected multivariate data analysis software. Finally, some existing research whose findings derive from unadjusted multi-sectoral samples may be invalid. Cross-sectoral variability does, therefore, appear to be a research problem in need of a general solution. It is disappointing then that not only have we not been able to locate any literature bearing directly on it; but also that two requests for comments, made on the Allstat Net forum earlier this year, failed even to attract replies.

5. 0. Geographical spread and variety of research coverage: Overall, these 18 studies embraced 13 different countries worldwide, and in individual coverage ranged downwards from the international, to national and regional. This wide geographical spread encompassed of course considerable variety in legal, economic and fiscal regimes. And, as noted (above), it includes some surveys based on judgemental/ convenience samples, as well as those about whose formats nothing is known. Two main criticisms may be made. First is that collectively this great variety constitutes a major obstacle to the synthesising of findings. Second, there is no indication that any of the three multinational studies (3, 4, 17) took steps to standardise, and thus to facilitate, the comparison of the sampling units they had drawn from different countries.

6. 0. Export performance and other measures. Specific problems of export measurement found were two. First, as indicated in Chapter1, it is generally agreed that more than one item is needed to capture the export performance concept. While 11 per cent of these studies were using one item only (3, 10), the rest had chosen two or more. However, as will be seen from the table, the mix of variables employed by the latter ran counter to desirable standardisation. The second concern related to other aspects of export performance measurement; in particular, two of the three operationalisation variables employed in one study appeared to be less than satisfactory; one was the use of a nominal scale with only 3 points to measure profitability, the other the setting of an export intensity value of 10 per cent as the lower limit for successful exporting. It is an undemanding target and a number of exporters in our sample survey at that level of activity were found to be passive, discontinuous exporters. [We set our lower level at 25 per cent].

7. 0. Survey populations and main sources of data: Moser's advice that the first step should always be to define the population, after which we need to be aware, in making exclusions, of the gap this opens up between target and survey populations *(Moser, C., et al, 1971, p53)* does not appear to have been followed by the great majority of subsample researchers. Only one explicitly defined its survey population (US E-Award winners). A further 20 (?) per cent seemed to imply, without discussion, that their sampling frames were adequate. The remainder were silent on this aspect.

7. 1. But all of these studies used as a basis for their surveys existing organised data from one or more of five main types of source. Most popular were trade directories (44%) followed by government/ official lists (22%), chamber of commerce lists (17%), specialist and export promotion lists (11%), and, finally, mainly respondents to an earlier survey (6%). Two studies (11%) had given no details. Overwhelmingly, studies had used one source, but one had used three (and has thus been counted thrice).

7. 2. However, the failure of the majority to identify their data sources by name precludes any proper assessment of their adequacy. The general absence of such precise identification and discussion must be of concern because of the potential or actual shortcomings of eg directories (usually created for other purposes) when employed as sampling frames, especially in terms of coverage, errors and wastage. For example, one of the group of 18 researchers mentioned that his (unnamed) specialist list had an error rate of at least 25 per cent. Again, the writer found that the UK Fashion Exports 2000 members' directory (which is known to have been used earlier as a sampling frame for a study of apparel exporters) included only 59 per cent of the exporters named in his own sampling frame. Probably all sampling frame sources should be identified precisely in the related research reports.

8. 0. <u>Selection of survey elements</u>. Almost four-fifths (78%) of the subsample researchers had garnered their survey elements in one of two ways: census or sample. The remainder (22%) did not indicate in any clear way how theirs had been chosen. Of those who did, sampling predominated, accounting for some 61 per cent of the total. In turn, samples fell into three subcategories: random (44%), judgemental (6%) and unspecified (11%). Censuses made up the remainder (17%), with some uncertainty about one of these. Thus, overall, for at least a third of the total there is unclarity about how their survey elements were chosen and for two-thirds of these there is no indication at all. Dharanaj (17) said nothing whatsoever about how his two samples had been selected (and, given his other omissions, his study is probably the lowest quality one of the 18). The procedures used by the other studies appear to have been adequate- though the one clearly judgemental sample would of course lack external validity.

9. 0. <u>Sector and product coverage</u>. Adequate comparisons of these studies would require, inter alia, comprehensive details of the sectors and products researched by each of them. But only partial details are available. Some 50 per cent of these studies failed to provide any information about the actual products covered and, overall, there is also a considerable shortfall in naming information for the sectors embraced, with 17 per cent failing to supply any details at all. Of the almost three-quarters of the subgroup (72%) which were multi-sector studies, some studies ranged across industrial and consumer goods, covering a hetereogeneous collection of products that were often produced and marketed under very different conditions.

9. 1. With adequate product data for only 50 per cent of all studies, there is scope for a high degree of unrecognised variability in the remainder of the sample. For instance, Wilson highlights the wide and fundamental differences that exist between consumer and industrial sectors and reminds us that these stem from the essential dissimilarities in the characteristics of industrial goods, industrial services, industrial markets and industrial buyers, which he then identifies (*A. Wilson, 1973, p2-14*). This is obvious when we consider that a study embracing unspecified industrial and consumer goods could range in its coverage, say, from makers of supertankers in the former to producers of brassieres in the latter. And indeed the Das study- No 5- shows how variegated the products can be. It extended across seven very dissimilar sectors, from spices and textiles through handlooms, handicrafts and leather to engineering goods plus chemical and allied products. Thus treating these sectors, willy nilly, as if they and

their firms were similar risks unrecognised and unquantified variability; and the problem is exacerbated (as it is in a large part of this group) when the actual sectors are not even identified.

9. 2.   Nor are these problems unique to our subsample; Leonidou reports similar sector variety, also with shortcomings in information, in his assessment of some 46 empirical studies, two-thirds of which were cross-sectional investigations of from 5 to 9 multiple industrial groups. He also comments on one aspect of the risks: namely that the 'spread of the research effort into a variety of industries may lead to biased findings since no account is taken of the potentially significant effects of various industry–specific factors on managerial characteristics' *(Leonidou, L.C., et al, 1998, p6).* (Their recognition that a narrower focus impedes generalisation is noted, but we have of course tried to show above why, without adjustments to data, such generalisation may be invalid). As no more than a starting point towards the reduction of inter-sector variability, it would appear desirable that future empirical studies standardise the use of SITCs as sector identifiers.

10. 0. Variability of firm size descriptions**:** The use of standardised definitions of enterprise size facilitate inter-study comparisons mainly   by enabling us to focus on the relevant differences between firms. But the firm size data, provided by 80 per cent of these researchers (the rest gave no, or no specific information), did not use a common descriptive standard. Overall, size information was variously reported as turnover, in a range of currency types, in terms of number of employees per firm, or in both forms. And, although the small, medium and large terminology was used by most, there was no common agreement on the definitions of these categories in terms of turnover and/ or number of employees (Table A1-1)

10. 1. These `March Hare' variations in the use of size terms is potentially very misleading. Thus it will be noted that Moini (7) sees small firms as those with fewer than 500 employees; Chan (3) regards this range as also including medium-sized firms; and Holzmuller(2) considers that small and medium firms comprise the much wider range of 50 to 1000 employees.   At best, the reader of research by these scholars would need to remind himself constantly what the writer intended to convey by the term eg 'medium**,** by referring to their size data, thus interposing another obstacle to understanding; at worst, there is unlikely to be enough information to reclassify into standard size categories, which, if achieved, might well then involve a mismatch /conflict between the text and the new size categories. Scholars need a common measuring rod with which to discuss and describe their findings. That used by the UK's Office for National Statistics, based on number of employees: Small (1-49), SME (50 - 249), Large (250+); ought to serve well. (ONS personal e-mail to writer, 2003). A final problem is that not all those defining size in turnover terms have given currency exchange rates.

11. 0. Nature and extent of follow-up of surveys**:** Unsurprisingly, and as scholars recognise, there is broad agreement in the literature that effective follow-up plays an important part in achieving an adequate response. It is especially necessary in the case of mail surveys. As Fowler puts it: '…there is no question that the most important difference between good mail surveys and poor mail surveys is the extent to which researchers make repeated contact with correspondents'. (Fowler, F.J., op cit., p46). Clearly, then, research studies ought to report on this aspect. But it is noteworthy that only 17 per cent reported following-up, with more than three-quarters (78%**)** giving no details of what follow-up if any, they had engaged in – and this group includes the studies that achieved the lowest reported response rates.

Comment [COMMENT1]:
1.46..* **Type Of Survey & Selection Of Survey Elements:** Three of these studies(1,8,12) conducted census surveys, 11 used samples, and one (14) gave no details of how it had chosen its survey elements, as indeed nothing about industry covered, no-response ratio **etc.?**). One study (1) arrived at its survey sample by whittling down its census firms, using seemingly appropriate rejection criteria. Of the other studies that employed samples,more than half (2,3,7,9,10, 11, 13) used random methods of selection. Three of the others (5, 14,15) used judgemental methods**(do they generalise from sample?)** , while the other three (1,4,6) gave no details whatsoever of how their samples had been chosen. T(*OPM: **How do ZS tabulate them***?*) Thus, for some 40 per cent of these studies key information needed to assess such aspects as the extent of sampling error (**do they give it**) is either not available or is not provided.

12. 0. Response rates: Response rate achieved is an important indicator of a study's quality. The lower the response rate in a survey the greater the risk of biased findings, especially when the response falls below 50 per cent *(Rodeghier, R., 1996, p38, 39)*. And Fowler points out that any response rate of, say, only 5 to 20 per cent, essentially a self-selecting sample, is very unlikely to provide any credible statistics about the characteristics of the population as a whole *(Fowler, F.J, 1993, p40)*. With reported response rates of less than 50 per cent for more than three-quarters (78%) of the subsample; and of less than 30 per cent for a third (including two of under 20 per cent), the response rates appear less than satisfactory- a conclusion again supported by expert opinion. '*...A poor response must constitute a dangerous failing and if it does not rise above, say, 20 or 30 per cent, the failing is so critical as to make the survey results of little if any value*'. (Moser, C., et al, 1971, p268). Nor are these low response rates exceptional. Thus Gemunden's review of 49 empirical export studies found that the typical study showed a response rate below 30 per cent (Gemunden, H.G., op cit, 1991, p43).

[Only two of the systematic studies achieved response rates of 50 per cent or above (Baldauf, Thirkell), with the others falling off quite steeply (Ogunmokun, 38%: Dharanaj 24% and 15%). Of the four, only one (Baldauf: 2 sets of phone calls) reported any follow-up of their mail shots. Moreover, none of the four gave details of any tests they had conducted to assess non-response bias; it may not unreasonably be concluded that no such tests were carried out]. **Note**: Bracketed results have yet to be integrated into 12.0, 13.0).

13. 0. Tests for non-response bias: Where response rates are low, as were many of our group of 18 studies, it is of course particularly important that tests for non-response bias be conducted. But only 28 per cent (9, 10, 11, 12, 16) reported conducting such tests and it was unclear whether another (No 7) had done so. On the reasonable assumption that here absence of evidence is indeed evidence of absence, the great majority of these studies (predominantly with rates below 50 per cent) did not conduct any appropriate tests. Moreover, none of those who did- even when response rates were as low as 9 and 4 per cent- reported any negative results (See Chapter 3 for a fuller discussion of these tests). Also noted was a disappointing tendency for those with poor response rates to seek to justify them by reference to other similarly inadequate results; failing to mention, however, that all concerned might well have had better outcomes had they conducted adequate follow-up. Perhaps the most cynical of these was the researcher who called in aid the Gemunden 1991 review, seeking to show that his own response rate was fairly typical, without bothering to add that Gemunden had been highly critical of low response rates and of the value of the studies based upon them.

14.0. Dates and timings: Much more than dates of publication (which may, however, have other implications), the dates of sampling frames and surveys can have an important bearing on the comparability and validity of empirical studies; in particular, it is important to report both the dates of any sampling frame used (eg business directory) and that of the related survey; the closer together they are clearly the better. Our sample of 18 studies falls far short of these desiderata. Publication dates and related details are of course available for all. But only one study (12) gave date data for both sampling frame and survey. Otherwise no information was provided about when surveys were conducted and only four other studies gave dates for their sampling frames. The intervals between the latter and publication year range from 3 to 5 years for three of them to an apparent 14 years for a fourth. Thus, for some two-thirds of these research projects nothing precise is known about when they were actually carried out. The first of two related problems is that the scope for comparing and contrasting accurately the studies is reduced, in that their relative positioning in time is not known and thus, eg whether they experienced the same or different trading conditions. Second, the wider the time interval between sampling frame and survey dates, the more erosion of the former

there will have been. For instance, for one study the dates were respectively 1992 and 1997. If, say, this survey was not conducted till 1996 its underlying sampling frame would inevitably have changed considerably in the meantime. By 2000, for example, our 1989 random sample of 30 knitting firms had lost 70 per cent of its original membership.

[Of the four systematic studies, only one (Thirkell,) provided details of the timing and duration of the field research]. **Note**: *For integration into 14.0.*

15. 0. Data analysis**:** In his review Zou had noted the diversity of methods used for data analysis and had gone on to comment on: the possible misuse of multiple regression; the probable need for path analysis and SEM; the hazards of applying multivariate analysis to small samples; the need to move beyond exploratory approaches; and the limited use of both theory in studies and of hypothesis testing.

15. 1. Our methodological assessment in this area had found, in addition to five exploratory approaches, these difficulties: four studies had used one or other of stepwise regression or discriminant analysis, an approach criticised by experts, particularly when used to `fish'; and although multivariate analysis is generally considered desirable, one study (Beamish) whose sample ought to have been large enough for m/v had restricted himself to bivariate analysis;

16. 0. Overall View Of Methodology;.This section of the thesis has broadly assessed the methodologies of a representative subsample of empirical export performance studies, more than half drawn from those included in Zou..S. et al's 1998 review. It summarised the shortcomings they had found, suggested broad criteria for sound methodology, recognised the latter's crucial importance to the reliability of study results, and noted that neither Chetty (1993) nor Zou (1998) had fully assessed the methodologies used in the empirical studies which they had respectively reviewed, nor had they mentioned Gemunden's useful 1991 review.

16. 1. Our assessment of the subsample appears to have found significant weaknesses overall, in the form of either or both lack of information and difficulties highlighted by what was actually disclosed. Twelve aspects of firms methodology that gave cause for concern were identified. Most of the researchers failed to define their survey populations. Sampling frames were almost universally existing lists, but only a few named these documents, or reported upon their suitability. Some 40 per cent failed to describe either the type of survey or the selection method used for its elements; some samples lacked external validity. There was a lack of uniformity in the scale of geographical coverage, as well as unclarity about how some surveys had been designed. The extent of variability cannot be assessed for the (more than half of) studies which did not report adequately or at all on the sectors and products surveyed. Lack of standardised firm size definitions is a significant obstacle to inter-study comparisons. Doubts about the reliability of the reported results for the (majority of) studies with below 50 per cent response rates are compounded by the general lack of information on response-related follow-up, if any, undertaken; these studies include those with the lowest response rates. And the majority failure to give dates for one or both of sampling frames and surveys is another source of potential error.

16. 2. The overall effect of the weaknesses found in these twelve areas**,** together with those arising from the **other reviews** reported above (including reviewers' use of the weakest form of meta-analysis), appears to cast considerable doubt on the reliability of the results reported in these various studies. This fact had important implications for the kind of empirical study which we undertook. In particular, it sought to make good as many of these failings as

Comment [COMMENT2]:
1.46..* **Type Of Survey & Selection Of Survey Elements:** Three of these studies(1,8,12) conducted census surveys, 11 used samples, and one (14) gave no details of how it had chosen its survey elements, as indeed nothing about industry covered, no-response ratio **etc.**?). One study (1) arrived at its survey sample by whittling down its census firms, using seemingly appropriate rejection criteria. Of the other studies that employed samples,more than half (2,3,7,9,10, 11, 13) used random methods of selection. Three of the others (5, 14,15) used judgemental methods**(do they generalise from sample?)** , while the other three (1,4,6) gave no details whatsoever of how their samples had been chosen. T(*OPM: **How do ZS tabulate them**?)* Thus, for some 40 per cent of these studies key information needed to assess such aspects as the extent of sampling error (**do they give it**) is either not available or is not provided.

possible, including focusing on one sector partly to avoid the problem of inter-sector variability.

Table A1-1   Subsample Of 18 Studies: Disparate Company Size Data Provided

Study No.	Company size data
5, 12	None
13	Non-specific: Small & m/sized firms chosen to be as variable as possible.

Turnover Data:

1	From less than $1m to $401m in 9 categories
4	Small and medium only: Canada, under $25m; in UK, under £12.5m.
11	Small $50k-500k; medium $500k-2,500k; large $2,500k-5,000k; v.large $5,000k+
16	Austrian Sch 20m+ (about *E*1.5m). Most of sample *E*7.3m - *E*36.3m. No small firms.
15	Small <NZ$ 10m (41%); Med: NZ$10-50m (30%);Larger >NZ$50m (29%).

Employee Numbers Data

2.	Small and medium firms from 50 to 1000 employees
3.	Small and medium firms with up to 500 employees
6	Aimed to include small and medium and selected firms with 5 to 200 employees
7	Small exporters with fewer than 500 employees
8	Employees: <100 (52%); 101-300 (27%); 300+ (22%)
9	Fewer than 500 employees; only 2.3% of sample exceeded 250 employees
10	550 exporting firms, each with 250 or more employees.
14	All but one of the firms had < 100 employees and a single line of business

Employee and Turnover Data

17	Employees: 1-500; Sales : <$1m to <$50m.
18	Employees: <20 =38%: 20 to 50= 20%: >50 =42% (of which 12% >200Es) Sales: A$5m or more =51%: <A$5m = 49%

Source: 18 study reports. Compiled by the writer

(Note: This Table also appears in Chapter 1- with a different table number)