Congestion in the Chinese automobile and textile industries revisited

A.T. Flegg and D.O. Allen[*]

Abstract

This paper re-examines a problem of congested inputs in the Chinese automobile and textile industries, which was identified by Cooper et al. (Socio-Economic Planning Sciences 35 (2001) 227–242). These authors employed a single approach to measuring congestion, however, so it is of interest to see whether other approaches would yield very different answers as regards the severity of this problem. Indeed, the measurement of congestion is an area where there has been much theoretical debate but relatively little empirical work. Here we use the data set assembled by Cooper et al. for the period 1981–1997 to compare and contrast the measurements of congestion generated by three alternative approaches. We find that these measurements are indeed very different.

Keywords: Data envelopment analysis; Congestion; Inefficiency

1. Introduction

In an interesting paper, Cooper et al. [3] examine the problem of inefficiency in the Chinese automobile and textile industries. Using annual data for the period 1981–1997, they find evidence of *congestion* in both industries. Congestion refers to a situation where the use of a particular input has increased by so much that output actually falls. In this sense, it can be viewed as an extreme form of technical inefficiency. Cooper et al. focus on the problems caused by the employment of excessive amounts of labour in these two industries and they discuss ways in which congestion could be managed without engaging in massive layoffs of workers. Here they demonstrate how output could be enhanced by improving managerial efficiency, while maintaining the size of the labour force.

Our aim in this paper is rather different. Instead of focusing on policy issues, we examine

---

[*] Tony.Flegg@uwe.ac.uk and David.Allen@uwe.ac.uk. School of Economics, Bristol Business School, University of the West of England, Coldharbour Lane, Bristol BS16 1QY.

the magnitude of the problem of congestion in these two industries and whether it makes much difference as to how we measure congestion. Although the theoretical issues surrounding the measurement of congestion have been discussed in several recent papers, no consensus has emerged regarding the most appropriate way to identify and measure congestion.[1] Indeed, the two main schools of thought − those associated with Färe and Grosskopf, on the one hand, and with Cooper et al., on the other − appear to be as divided as ever.[2] What is more, the theoretical discussions have focused, to a large extent, on a relatively narrow issue: whether congestion does or does not exist in a particular case, as opposed to *how much* congestion there is likely to be. Unfortunately, apart from the earlier study by Cooper et al. [9] and the work on British universities by Flegg and Allen [17], there is very little published evidence available to offer guidance as to whether the competing approaches are apt to yield very different results in reality. Our aim is to augment this limited stock of empirical evidence.

The rest of the paper is structured as follows. We begin by explaining Färe and Grosskopf's radial approach to the measurement of congestion. This is followed by a discussion of the slacks-based procedure of Cooper et al. We then use the data set assembled by Cooper et al. for the Chinese automobile and textile industries to compare and contrast the measurements of congestion generated by the two approaches. We also consider some results from a new approach proposed by Tone and Sahoo [20]. Finally, we present our conclusions.

2. Färe and Grosskopf's approach

This axiomatic approach to the measurement of congestion was given operational form in a monograph published in 1985 by Färe et al. [14]. This classic study gave rise to numerous applications. For ease of exposition, this procedure is referred to hereafter as Färe's approach. A big advantage of Färe's approach is that is possible to decompose his measure

---

[1] See, for example, [2], [6], [7], [9], [10] and [12].

[2] The earlier criticisms of Färe and Grosskopf's approach by Cooper and his associates are reiterated in Cooper et al. [5], who remark (p. 189): "Undoubtedly these defects in [this] approach can be eliminated or alleviated but the research needed to do this has not yet been done."

of overall technical efficiency (TE) in a straightforward way into pure technical efficiency (PTE), scale efficiency (SE) and congestion efficiency (CE), using the identity:

$$TE \equiv PTE \times SE \times CE, \tag{1}$$

where TE = 1 and TE < 1 represent technical efficiency and inefficiency, respectively.

Figure 1 near here

To illustrate the use of Färe's approach, consider Figure 1. This shows six decision-making units (DMUs), labelled A to F, each using a different combination of two inputs, x_1 and x_2, to produce an output of y = 1. This example assumes constant returns to scale, so that SE = 1, and makes use of an *input-oriented* approach. DMUs C and D are clearly technically efficient, whereas E is inefficient. In terms of identity (1) above, TE = PTE = 0.5 for E. The status of the remaining DMUs is less straightforward to determine, as it depends on one's assumptions regarding the underlying technology. In particular, we need to distinguish between *strong* and *weak* disposability.

Strong (or *free*) disposability occurs when the slack in a particular input can be disposed of at no opportunity cost. In Figure 1, the boundary S_1CDS_2 defines the strongly disposable isoquant for y = 1, so that all quantities of x_1 and x_2 in excess of 2 are assumed to be freely disposable. For instance, a rise in either x_1 or x_2 from, say, 2 to 3 would not reduce output. Thus neither x_1 nor x_2 would exhibit congestion. By contrast, under weak disposability, an equiproportionate increase in *both* inputs is assumed not to reduce output. The boundary W_1ACDFW_2 defines the weakly disposable isoquant for y = 1.

Weak disposability allows for the occurrence of upward-sloping isoquant segments such as CA and DF in Figure 1. Such segments require the marginal product of one of the inputs to be negative, thus enabling congestion to occur.[3] In the case of DF, it is reasonable to suppose that x_1 is the input with a negative marginal product. Hence output would remain constant at y = 1 along DF because a simultaneous rise in the quantities of both inputs would

---

[3] Note that the gradient of each line segment, dx_2/dx_1, equals $-MP_1/MP_2$. These marginal products must have opposite signs to yield $dx_2/dx_1 > 0$.

3

cause exactly offsetting changes in output (a rise due to x_2 and a fall due to x_1). It should be noted that the axiom of weak disposability precludes the possibility of both inputs having negative marginal products.[4]

We can now proceed to classify the remaining DMUs in terms of identity (1) above. With respect to A, Färe's analysis might proceed as follows. Because A is on the weakly disposable isoquant for $y = 1$, it would be free from *pure* technical inefficiency (PTE = 1), yet this DMU would be held to be suffering from congestion. Its CE score, as measured by the ratio OA′/OA, would equal ⅔. Its TE score would also equal ⅔, the product of PTE = 1 and CE = ⅔. Congestion would arise owing to the difference between the upward-sloping isoquant segment CA, which is assumed to exhibit weak disposability, and the hypothetical vertical dashed line emanating from C, which is characterized by strong disposability. If A were able to move to point A′, and thereby get rid of its congestion, it could attain TE = 1. Likewise, F would need to move to point F′ in order to eradicate its congestion. Its TE score would also equal ⅔, the product of PTE = 1 and CE = ⅔. In contrast, B would exhibit both pure technical inefficiency and congestion: PTE = OB′′/OB ≈ 0.56 and CE = OB′/OB′′ ≈ 0.89, so that TE = 0.5 ≈ 0.56 × 0.89.[5]

It is worth noting that negative marginal productivity is necessary but not sufficient for congestion to occur under Färe's approach. To demonstrate this point, suppose that we removed F from the data set. The weakly disposable isoquant for $y = 1$ would then pass through E, yet this DMU would not exhibit congestion. Thus an upward-sloping isoquant (which requires the marginal product for one of the inputs to be negative) does not entail congestion under Färe's approach. In fact, for congestion to be identified, the relevant ray would need to cross the horizontal dashed line emanating from D. This condition is satisfied in the case of F but it is not satisfied with respect to E.

---

[4] For more discussion of the issues surrounding strong and weak disposability, see [19, pp. 170, 175–186] and [21, pp. 181–186].

[5] The calculations were carried out using *OnFront* (www.emq.com).

It is also worth mentioning that the presence of slack is necessary but not sufficient for congestion to occur under Färe's approach.[6] This point is illustrated by the fact that a congested DMU like F has slack of DF′, whereas an uncongested DMU like E has no slack. In addition, any DMU situated at a point such as g or h in Figure 1 would not be regarded as being congested, despite the presence of a substantial amount of slack in the relevant input.

3. Cooper's approach

This alternative approach to the measurement of congestion was first proposed in 1996 by Cooper et al. [11]. It was subsequently refined by Brockett et al. [1] and by Cooper et al. [9]. For simplicity, this procedure is referred to hereafter as Cooper's approach. The analysis, as described here, proceeds in two stages.[7] In the first stage, the *output-oriented* variant of the well-known BCC model is employed to obtain an efficiency score, ϕ^*, for each DMU, along with the associated BCC input slacks.[8] In the second stage, the amount (if any) of slack in each input that is associated with technical inefficiency (as opposed to congestion) is identified. This then makes it possible to measure the amount of congestion as a residual. The relevant models are specified formally in [3], so the details need not detain us here. Instead, we present a graphical illustration to highlight the salient points.

Before considering this illustration, however, we need to define Cooper's measure of congestion, which is denoted here by C_C. The first step is to specify a formula for calculating the amount of congestion:

$$c_i = s_i^* - \delta_i^*, \tag{2}$$

where c_i is the amount of congestion associated with input i, s_i^* is the total amount of slack in input i and δ_i^* is the amount of slack attributable to technical inefficiency. The asterisks

---

[6] See [12] and [5, pp. 184−189].

[7] It is possible to combine these two stages but this would entail sacrificing some valuable information in terms of the results provided. See [4].

[8] Cooper et al. [10, p. 211] state that it would be wrong to use an input-oriented model when implementing their approach.

denote optimal values generated by the DEA software. The measured amount of congestion is thus a residual derived from the DEA results. We can then rewrite equation (2) as follows:

$$c_i/x_i = s_i^*/x_i - \delta_i^*/x_i, \tag{3}$$

where c_i/x_i is the proportion of congestion in input i, s_i^*/x_i is the proportion of slack in input i and δ_i^*/x_i is the proportion of technical inefficiency in input i. The final step is to take arithmetic means over all inputs to get:

$$\overline{C}_c = \overline{s/x} - \overline{\delta/x}. \tag{4}$$

Hence $\overline{C}_c$ measures the average proportion of congestion in the inputs used by a particular DMU. It has the property $0 \le \overline{C}_c \le 1$. Cf. [9, p. 11].

Figure 2 near here

Now consider Figure 2, which portrays the situation facing nine DMUs. The BCC model, in the context of a simplified production function $y = f(x)$, is depicted by the convex VRS (variable returns to scale) frontier ABCDEF and its horizontal extension from F. The diagram also shows, for comparison, the linear CRS (constant returns to scale) frontier obtained from the CCR model, which is produced if we drop the convexity constraint.[9] The issue of congestion arises from the inclusion of G in the diagram.

Proceeding clockwise around the diagram, we note that DMUs A to E would have $\phi^* = 1$ and $s^* = 0$. Hence these DMUs would be BCC efficient and thus uncongested. F would have $\phi^* = 1$ but $s^* = 1$. It would be classified as being BCC inefficient, yet uncongested. This is because the elimination of its slack of one unit would not alter output. In terms of formula (2) above, we would have $c = 1 - 1 = 0$. By contrast, with $\phi^* = 1.25$ and $s^* = 2$, G would be classified as being congested. This is because of the potential increase in output from $y = 4$ to $y = 5$. However, only half of its total slack of $s^* = 2$ would represent congestion; the remainder would be classified as technical inefficiency. Using formula (2) above, we would

---

[9] See [8] for a detailed discussion of the CCR and BCC models.

have c = 2 − 1 = 1, i.e. one unit of congestion, while the proportional formula (3) above

would yield c/x = 2/8 − 1/8 = 0.125.

Because H and I are located beneath the frontier, there is clearly scope for a rise in output

(in fact, $\phi^* = 1.25$ in both cases), yet neither DMU would be deemed to be congested under

Cooper's approach. This is because output cannot be augmented by the elimination of slack:

I has s* = 0 and thus cannot be congested, whereas H has c = 1 − 1 = 0.

The aim of the second stage of Cooper's analysis is to ensure that congestion is not

confused with technical inefficiency. The crucial point here is that slack only represents

congestion if there is a potential rise in output. However, this second stage is redundant if all

DMUs with scores of $\phi^* = 1$ are located at extreme points on the BCC frontier. This means

that the data set cannot contain either (i) weakly efficient DMUs or (ii) efficient DMUs that

can be expressed as weighted averages of other efficient DMUs. In Figure 2, these cases are

exemplified by DMUs F and C. In reality, these conditions are not too difficult to satisfy,

which means that computing Cooper's measure of congestion is much more straightforward.

This issue is taken up later in the paper.

4. The data set

Before presenting the DEA results, it is worth examining the salient features of the data

set assembled by Cooper et al. [3] for the Chinese automobile and textile industries in the period

1981−1997. The variables, which are graphed in Figures 3 and 4, are defined as follows:

Y is output measured in units of 1 million renminbi, in 1991 prices;

K is capital measured in units of 1 million renminbi, in 1991 prices;

L is labour measured in units of 1000 persons.

Figures 3 & 4 near here

What is most striking about Figures 3 and 4 is the extent to which output has grown in

both industries, especially since 1990. As noted in [3], that year marked the end of the

Chinese government's policy of focusing on maintaining employment in Chinese industries, and the introduction of a new market-oriented policy aimed at enhancing efficiency. Figures 3 and 4 also reveal marked differences in the behaviour of the inputs. For instance, whereas employment in automobiles rose fairly steadily throughout the period, employment in textiles first rose and then fell from 1991 onwards. The behaviour of capital is also very different in the two industries: hardly changing in textiles, yet growing strongly in automobiles. Indeed, it is noticeable how closely the variables Y and K track each other in automobiles.

5. Results from Cooper's approach

In this section, we build upon the work reported in [3], by providing some additional results and analysis. In particular, we examine the overall technical efficiency, as well as scale efficiency, of each industry and, to facilitate comparisons, we express congestion in relative terms. Our findings are displayed in Tables 1 and 2. The textile industry will be considered first.

Tables 1 & 2 near here

The CCR efficiency scores in Table 1, which measure overall technical efficiency (TE), show that the Chinese textile industry was highly technically inefficient in the 1980s. For instance, in 1981, it was producing only 27.6% of its potential output. However, the TE scores improved dramatically in the 1990s, with TE = 1 being achieved by 1996. Table 1 also reveals that scale inefficiency was either wholly or partly responsible for the very low TE scores in the period up to 1985. Thereafter, scale efficiency (SE) improved considerably and this was an important factor behind the impressive rise in overall technical efficiency that took place in the textile industry during the period as a whole.[10] In fact, this industry exhibited increasing returns to scale up to 1987, which suggests that it was operating at too small a scale to be efficient. Thereafter, returns to scale were sometimes decreasing and

---

[10] SE is calculated by dividing the CCR score by the BCC score. The latter measures technical efficiency once the effects (if any) of an inappropriate scale have been removed.

sometimes increasing, culminating in constant returns in the last two years.

In their abstract, Cooper et al. [3] describe the Chinese textile industry as being "heavily congested". However, Table 1 shows that labour congestion was confined to the period 1988 to 1992, with 1991 being an uncongested anomaly. The fact that the employment graph in Figure 3 exhibits a pronounced "hump" during this period lends support to this finding. Table 1 also points to a limited amount of capital congestion in 1982 and 1989, along with a much more serious problem in 1995.[11]

The results displayed in Table 2 for automobiles offer an interesting contrast with those for textiles. Most striking is the fact that there are eleven instances of labour congestion in automobiles, as opposed to only four in textiles. What is more, where both industries are congested in this way, the problem in relative terms is much more severe in automobiles.[12] A comparison of the employment graphs in Figures 3 and 4 suggests a possible explanation: whereas the labour force in automobiles rose fairly steadily during the period under review, that in textiles first rose and then fell.

Another contrasting feature is the fact that scale efficiency in automobiles was relatively high throughout the period, whereas textiles experienced severe scale inefficiency in the early years. In addition, apart from the period 1986 to 1988, the TE scores for automobiles are mostly fairly high and do not exhibit any obvious trend. By contrast, overall technical efficiency in textiles rose substantially during the period under review.

Also worth noting is the relatively high (and similar) capital congestion in the two industries in 1995. In the case of automobiles, Cooper et al. [3, p. 235] suggest that this might have been due to overinvestment by the Chinese government and foreign investors. However, apart from 1995, Table 2 indicates that capital congestion was not a problem in the automobile industry during the period under review.

---

[11] Cooper et al. [3] miss the capital congestion in textiles in 1989, owing to an error in their Table 2, which shows no slack for that year (the slack was actually 0.52).

[12] In absolute terms, the differences are much less marked. This is due to the substantially larger labour force in textiles. See [3, Tables 1 and 2].

To complete this discussion of Cooper's approach, a key methodological issue needs to be considered. As noted earlier, this approach proceeds in two stages. In the first stage, the BCC input slacks are determined. In the second stage, the amount (if any) of each slack that is deemed to represent technical inefficiency is calculated. What remains of this slack is then assumed to represent congestion.

However, if all DMUs on the BCC frontier are located at extreme points, this second stage is redundant, so it is of interest to see whether this is the case here.[13] Even though no weakly efficient DMU (year) was identified, i.e. one with $\phi^* = 1$ but non-zero input slack, three instances (out of 34 possible cases) were found where a particular year could be expressed as a weighted average of other years. The years in question were 1983 and 1984 for textiles, and 1982 for automobiles.[14] As expected, deleting each of these years in turn from the data set did not alter the values of ϕ^* for the remaining years. Hence these years are akin to DMU C in Figure 2.

Whilst the second stage of Cooper's model is not redundant in this instance, its use makes little practical difference to the results. Only three of the labour congestion scores were changed as a result of omitting the second stage, and then only slightly, and the capital congestion scores were unaffected.[15] Therefore, in view of the extra computation required, we would question whether the benefits in terms of additional accuracy from using this refinement are worth pursuing in practical applications. By contrast, the stage 1 results can easily be generated via a standard DEA software package. We used *DEA-Solver Pro* (www.saitech-inc.com) to generate the slacks and Excel to perform the calculations. We then adjusted these stage 1 results to take account of any technical inefficiency components, by

---

[13] This issue is discussed in Cooper et al. [9, pp. 15–17], who remark (p. 15) that "the frontiers … in most real world data sets contain only the extreme efficient DMUs".

[14] For example, $Y_{82} = \alpha Y_{81} + (1 - \alpha)Y_{83}$, $\alpha \approx 0.755$. Such cases can be identified by the fact that they do not have any other DMUs in their BCC reference sets.

[15] Incorporating the second stage had the effect of lowering the labour congestion score from 0.003 to zero for textiles in 1993; for automobiles, this score fell from 0.224 to 0.219 in 1990, and from 0.008 to zero in 1996.

making use of the values of δ_i^* reported by Cooper et al. [3, Tables 2 and 3].

6. Results from Färe's approach

Before examining the results from this approach, we need to take some recent theoretical developments into account. Two key issues need to be considered. The first concerns the *order* in which overall technical efficiency (TE) is decomposed into pure technical efficiency (PTE), scale efficiency (SE) and congestion efficiency (CE). Here it is conventional in applied studies to postulate strong disposability when measuring scale effects, and only then to allow for the possibility of congestion.[16] This is also the approach underlying Cooper's procedure. However, Färe and Grosskopf [13] have highlighted the difficulty of discriminating between scale inefficiency and congestion; they point out that the CE score will depend on the order in which TE is decomposed.[17] Their solution is to specify CRS rather than VRS technology in situations where congestion can be anticipated on *a priori* grounds.

The second issue concerns the *orientation* of the model and the distinction between input and output congestion. In the current version of *OnFront*, the software supporting Färe's approach (www.emq.com), congestion of inputs is measured using an input-oriented approach, whereas congestion of outputs is captured via an output-oriented approach.[18] With regard to outputs, congestion refers to a situation where one or more of the outputs is an undesirable by-product of joint production; for instance, air pollution associated with the generation of electricity [15].

In accordance with the above arguments, we have employed an input-oriented variant of Färe's approach, with CRS as the underlying technology, to compute the CE scores. This approach is consistent with the earlier discussion surrounding Figure 1. However, for

---

[16] For a recent example, see [18].

[17] In the identity TE ≡ PTE × SE × CE, TE and the product SE × CE are unaffected by the order of the decomposition but the individual values of SE and CE are affected.

[18] We are grateful to Pontus Roos, of the Institute of Applied Economics in Sweden, for clarifying this issue for us.

comparison, we have also computed these scores using the older procedure based on VRS. Our findings are reported in Tables 3 and 4. The textile industry will again be considered first.

Tables 3 & 4 near here

Whilst the use of an input orientation does not affect the TE scores, it does generate a new set of VRS-based SE scores for the textile industry, which are shown in Table 3. With one exception, these scores are either the same or noticeably lower than those in Table 1 and thus indicate more scale inefficiency in many cases than was suggested by Cooper's method.

Switching from VRS to CRS has the effect of raising the SE scores in several years. However, the most striking impact of the change in technology is on the CE scores: if we posit VRS, congestion appears only twice, whereas it is evident in all but five years under CRS! To shed some light on this interesting finding, we carried out some further analysis to determine which input was responsible for the congestion in each congested year. This involved running the model thrice, each time under a different pair of assumptions about the disposability of the two inputs: (strong, strong), followed by (strong, weak) and then (weak, strong). Whereas strong disposability precludes congestion, weak disposability makes it possible to identify congestion if it exists.[19]

In Table 3, the input responsible for congestion is indicated by the L or K appended to the CE score. Here it is interesting to see that the CRS-based results show a continuous period of labour congestion between 1985 and 1992. This is in contrast with the much shorter and non-continuous congested period from 1988 to 1992 that was indicated by Cooper's approach. Looking again at Figure 3, a period of labour congestion from 1985 to 1992 seems perfectly plausible. On the basis of these findings, it would be fair to describe the Chinese textile industry as being congested for a substantial part of the period under review but whether it would be accurate to describe it as being "heavily congested" is a moot point.

---

[19] To illustrate, suppose that we were checking for capital congestion. Model 1 would specify strong disposability for both K and L, whereas model 2 would specify strong disposability for L but weak for K. If model 2 then gave a higher efficiency score, this would indicate that K was indeed congested. See [19, p. 183] for a discussion of this approach.

The CRS-based results in Table 3 also reveal the presence of chronic capital congestion in 1981 and 1982. These early years are also interesting in the sense that they give very different outcomes for the relative amounts of congestion and scale inefficiency, depending on whether one assumes CRS or VRS. However, in the case of 1995, the CE scores are very similar; they are also consistent with the result shown in Table 1 for Cooper's approach.

The results for pure technical efficiency are also worth noting. There are seven cases of PTE < 1. Moreover, for six of these years, the CRS-based results show CE < 1 as well. The exceptional year, 1993, is akin to DMU E in Figure 1, i.e. located in the interior of the weakly disposable isoquant, yet uncongested. Table 3 indicates that, by eliminating pure technical inefficiency, the same output could have been produced with only 83.6% of the capital and labour actually used in 1993. The other six years, which have PTE < 1 and CE < 1, are akin to DMU B in Figure 1. Here the results show that, by eliminating both congestion and pure technical inefficiency, the same output could have been produced with considerably less capital and labour. Pure technical inefficiency is especially severe in 1988 and 1989.

Table 4 displays the findings for the automobile industry. What is most striking about these results is how little labour congestion they indicate as compared with the findings recorded in Table 2 for Cooper's approach. In particular, whereas Cooper's approach suggests that Chinese automobile industry was heavily congested in terms of labour in the period 1984–1993, Färe's CRS-based approach indicates no labour congestion whatsoever! Furthermore, even in those years where the VRS-based estimates signify the presence of labour congestion, the amounts involved are, in most cases, insubstantial.

Once again, the first two years generate some unusual results. If we posit VRS, then it is immaterial whether we follow Cooper or Färe: both approaches ascribe the overall technical inefficiency in 1981 and 1982 entirely to the existence of increasing returns to scale. By contrast, if CRS is assumed, then labour congestion is identified as the underlying problem! With the exception of those two years, however, Table 4 reveals that switching from VRS to

CRS has relatively little impact on the SE scores. These scores are generally high, which suggests that scale inefficiency was not a serious problem in the automobile industry. The same inference can also be drawn from Cooper's approach.

As regards the relatively low TE scores in 1986, 1987 and 1988, Table 4 shows that this inefficiency can largely be attributed to the incidence of pure technical inefficiency in those years. By contrast, Cooper's approach points to a substantial amount of labour congestion. It may be noted, finally, that capital congestion does not appear to have been a significant problem in the automobile industry, except perhaps in 1995.

7. Graphical comparison of results

The estimates of congestion generated by the two approaches are graphed in Figures 5 and 6. To facilitate comparisons, the congestion efficiency scores recorded in Tables 3 and 4 were re-expressed as inefficiency scores by subtracting them from one.

Figures 5 & 6 near here

Figure 5 illustrates the fact that the different measures yield markedly different outcomes with respect to the amount of labour congestion in the textile industry. Färe's CRS-based measure, $C_{F,CRS}$, points to a fairly long and continuous period of labour congestion from 1985 to 1992, yet Cooper's measure, C_C, identifies congestion in only four of these years and Färe's VRS-based measure, $C_{F,VRS}$, finds no congestion at all!

Even more striking differences are apparent in the automobile industry: based on the values of C_C, a lengthy period of severe labour congestion is evident, yet this finding is not corroborated by the other two measures. $C_{F,VRS}$ suggests that labour congestion was much less severe in terms of both magnitude and duration, while $C_{F,CRS}$ indicates no labour congestion apart from that in the first two years.

The results for the automobile industry are interesting in the sense that its labour force rose fairly steadily during the period under review and Cooper's measure suggests that this

14

growth in employment resulted in severe congestion. A very different picture emerges, however, if we use either of Färe's measures.

8. Congestion and diseconomies of scale

Tone and Sahoo [20] have proposed a new unified approach to measuring congestion and scale economies, which can shed some new light on the problem of congestion in the two industries. For simplicity, this procedure is referred to hereafter as Tone's approach. This approach is similar to that of Cooper inasmuch as a BCC output-oriented model is used initially, yet Tone measures congestion very differently. In particular, he uses a slacks-based measure in projecting congested DMUs onto the BCC frontier.

The analysis can easily be carried out using the *DEA-Solver Pro* software, which generates a congestion score, $\psi \geq 1$, for each DMU, along with a *scale diseconomies* parameter, $\rho < 0$, for each congested DMU. The results for the two industries are presented in Table 5.

Table 5 near here

In terms of identifying whether or not congestion exists, it makes little difference whether one uses Tone's approach or that of Cooper. The only exceptions are textiles in 1993 and automobiles in 1996. These years have congestion scores of 1.002 and 1.001, respectively, whereas $\psi = 1$ would be needed for full consistency between the two approaches. The discrepancies can be explained by the minute amounts of technical inefficiency identified in the second stage of Cooper's approach, which are ignored in Tone's procedure.

Where congestion does occur, Tone's approach complements that of Cooper by affording some useful additional information. Consider the findings for textiles. The congestion scores for this industry suggest that its output in certain years would have been far higher if congestion had not been present; for instance, by 29.1% in 1988, 134.6% in 1989 and 31.5% in 1990. The exceptionally large congestion score in 1989 reflects the fact that both factors were congesting in that year. By contrast, labour was the sole congesting factor in 1988,

1990 and 1992, while capital was the only source of congestion in 1982 and 1995.

The values of ρ are also worth examining. These measure the responsiveness of output to a greater use of congested inputs. For instance, for 1989, ρ is calculated as follows:

$$\rho = \frac{\% \text{ change in Y}}{\text{mean} \% \text{ change in K and L}} = \frac{-134.64\%}{0.5(2.0023\% + 6.1141\%)} = -33.18 \tag{5}$$

This strikingly large scale diseconomy suggests that an excessive use of both capital and labour in 1989 caused a very large loss of output. This evaluation is relative to 1994.

Where only one input is congested, ρ is simply the ratio of the percentage change in Y to the percentage change in the input in question. Here it is worth noting that $\rho = -14.55$ in 1982 but only -0.50 in 1995. These values suggest that the overuse of capital in the textile industry had a much bigger impact in 1982 than in 1995, yet Cooper's slacks-based measure (see Table 1) indicates substantially more capital congestion in 1995!

Table 5 suggests that congestion was more protracted for automobiles than for textiles. This confirms the outcome from Cooper's approach. Nonetheless, the congestion scores for automobiles in 1988, 1989 and 1990 are noticeably lower than the corresponding scores for textiles. This reflects the fact that the BCC efficiency scores are typically much lower for textiles than for automobiles (see Tables 1 and 2). These BCC scores put an upper limit on the possible size of the congestion scores under Tone's approach.[20] This factor can also explain why most of the values of ρ are relatively small for the automobile industry.

9. Conclusion

The focus of this paper has been on congestion in the Chinese automobile and textile industries, which was identified as a problem by Cooper et al. [3]. We have used their annual data for the period 1981−1997 to reassess the magnitude of this problem. As well as pursuing the approach advocated by Cooper et al., we have also employed two alternative procedures for measuring congestion: the conventional approach of Färe and Grosskopf, and a new method

---

[20] The congestion scores cannot exceed the reciprocals of the BCC scores shown in Tables 1 and 2.

proposed recently by Tone and Sahoo. In addition, two alternative versions of Färe and Grosskopf's approach were examined: one based on variable returns to scale (VRS) and the other on constant returns to scale (CRS). For ease of exposition, these alternative procedures are referred to here as the approaches of Cooper, Färe and Tone.

The Chinese textile industry is interesting in the sense that its labour force rose rapidly during the 1980s, before reaching a peak in 1991 and then declining. This shedding of labour was associated with the economic reforms introduced in 1990. Indeed, by 1997, employment had fallen back to the level prevailing in 1986. Cooper's procedure suggested that the rapid growth in employment in the 1980s led to appreciable labour congestion in the period 1988–1990, as well as in 1992. This impression was reinforced by the results from Färe's CRS-based measure, $C_{F,CRS}$, which pointed to the presence of considerable labour congestion lasting from 1985 to 1992. A very different picture emerged, however, when we used Färe's VRS-based measure, $C_{F,VRS}$: this indicated no labour congestion at all!

In contrast to what happened in textiles, employment in the automobile industry grew fairly steadily throughout the period under review. Cooper's approach indicated that this steady expansion was accompanied by a lengthy period of severe labour congestion, yet this finding was not corroborated by the other two measures. $C_{F,VRS}$ suggested that labour congestion was much less severe in terms of both magnitude and duration, while $C_{F,CRS}$ indicated no labour congestion apart from that in the first two years.

This marked difference in the findings from the alternative methods is rather disturbing. However, if we follow Färe and Grosskopf's suggestion [13] that one should specify CRS technology in situations where congestion is anticipated on *a priori* grounds, then this does limit the range of possible outcomes. In the case of textiles, we then have a choice between (i) the fairly short and interrupted period of labour congestion indicated by Cooper's method and (ii) the rather longer and continuous period suggested by Färe's procedure. For automobiles, the choice is starker: labour congestion confined to the first two years (Färe) or

an extensive period of severe labour congestion (Cooper).

One way of resolving the dilemma of which set of estimates to accept would be to choose the method of estimation on the basis of its theoretical properties. Unfortunately, this is complicated by the fact that the competing methods have particular merits and demerits. Here it is worth noting the observation by Cherchye et al. [2, pp. 77−78] that the original purpose of Färe's procedure was not to measure the amount of congestion *per se* but instead to measure the impact, if any, of congestion on the overall efficiency of a particular DMU. By contrast, Cooper's method was specifically designed to measure congestion. Hence the underlying aim of one's study should help to determine which method to choose. It was also noted earlier that negative marginal productivity is necessary but not sufficient for congestion to occur under Färe's approach. This is not a property of Cooper's method, so this could be a reason for using his procedure rather than Färe's. Two final considerations might be (i) whether an input or output orientation is thought to be most appropriate and (ii) whether one wishes to posit CRS or VRS as the underlying technology when measuring congestion.[21]

With minor exceptions, the approaches of Tone and Cooper generated the same sets of congested years. This outcome was anticipated because both procedures use an output-oriented VRS model as their point of departure. Tone's approach also gave estimates of the potential rise in output from eliminating congestion. The scope for raising output in this way was particularly noticeable in textiles, with potential rises of 29.1% in 1988, 134.6% in 1989 and 31.5% in 1990. The exceptionally large congestion score of 2.346 in 1989 can partly be explained by the fact that both factors were congested in that year. There were only two other cases of capital congestion in textiles and only one in automobiles. With this single exception, labour was the sole congesting factor in automobiles.

The results revealed that the textile industry suffered from chronic scale inefficiency up to 1987, yet its performance was much better thereafter. By contrast, scale efficiency in

---

[21] For a more detailed discussion of the pros and cons of the different approaches, see [16].

automobiles was generally high throughout the period under review. Indeed, textiles was the industry that made the most progress in terms of reducing all types of technical inefficiency.

It is worth attempting to reconcile the results obtained here with the earlier findings of Cooper et al. [9], who examined data for three Chinese industries (textiles, chemicals and metallurgy) over the period 1966−1988. In most cases, Cooper et al. obtained noticeably larger amounts of congestion when their own method was employed, as opposed to that of Färe and Grosskopf [9, Tables 2−4].[22] However, it should be borne in mind that, when computing Färe's measures, Cooper *et al.* assumed VRS rather than CRS.

If we focus on labour congestion and assume VRS, then it is certainly true that, in both industries and in all years, $C_C \geq C_{F,VRS}$, where C_C denotes Cooper's measure of congestion. This can be verified from Figures 4 and 5, bearing in mind that Färe's VRS-based measure registered no labour congestion in textiles in any year. Hence these results are consistent with those of Cooper et al. What is more, if we ignore the first two years, then $C_C \geq C_{F,CRS}$ is demonstrably true in the automobile industry. However, the ranking in textiles is more ambiguous: $C_C > C_{F,CRS}$ in 1988 and 1989 but not in any other year. Also, in most of these other years, $C_{F,CRS}$ exceeds C_C by a wide margin.

An important methodological finding of this paper is that is that it appears unnecessary, from a practical point of view, to use the second stage of Cooper's method. The minimal improvements in accuracy from using this refinement do not, in our opinion, justify the extra computational work involved.[23] Indeed, basing the calculations solely on the first stage would considerably enhance the attractiveness of Cooper's method to practitioners.

In view of the marked differences in the results generated by the alternative approaches examined in this paper, we believe that it would be wise for analysts not to rely on the outcomes from a single method of measuring congestion. We would also urge that careful

---

[22] We should note the observation by Färe and Grosskopf [12, pp. 32−33] that, as compared with Cooper's method, their own procedure would tend to yield a smaller amount of congestion.

[23] Indeed, Cooper et al. [9] found that using the second stage made no difference to their results.

attention be paid to the theoretical properties of the measures used.

References

1. P.L. Brockett, W.W. Cooper, H.C. Shin, Y. Wang. Inefficiency and congestion in Chinese production before and after the 1978 economic reforms. *Socio-Economic Planning Sciences* **32**, 1–20 (1998).

2. L. Cherchye, T. Kuosmanen, T. Post. Alternative treatments of congestion in DEA: a rejoinder to Cooper, Gu, and Li. *European Journal of Operational Research* **132**, 75–80 (2001).

3. W.W. Cooper, H. Deng, B. Gu, S. Li, R.M. Thrall. Using DEA to improve the management of congestion in Chinese industries (1981–1997). *Socio-Economic Planning Sciences* **35**, 227–242 (2001).

4. W.W. Cooper, H. Deng, Z.M. Huang, S.X. Li. A one-model approach to congestion in data envelopment analysis. *Socio-Economic Planning Sciences* **36**, 231–238 (2002).

5. W.W. Cooper, H. Deng, L.M. Seiford, J. Zhu. Congestion: its identification and management with DEA. In *Handbook of Data Envelopment Analysis* (Edited by W.W. Cooper, L.M. Seiford, J. Zhu), pp. 177–201. Kluwer, Boston (2004).

6. W.W. Cooper, B. Gu, S. Li. Comparisons and evaluations of alternative approaches to the treatment of congestion in DEA. *European Journal of Operational Research* **132**, 62–74 (2001).

7. W.W. Cooper, B. Gu, S. Li. Note: alternative treatments of congestion in DEA – a response to the Cherchye, Kuosmanen and Post critique, *European Journal of Operational Research* **132**, 81–87 (2001).

8. W.W. Cooper, L.M. Seiford, K. Tone. *Data Envelopment Analysis*. Kluwer, Boston (2000).

9. W.W. Cooper, L.M. Seiford, J. Zhu. A unified additive model approach for evaluating inefficiency and congestion with associated measures in DEA. *Socio-Economic Planning Sciences* **34**, 1–25 (2000).

10. W.W. Cooper, L.M. Seiford, J. Zhu. Slacks and congestion: response to a comment by R. Färe and S. Grosskopf. *Socio-Economic Planning Sciences* **35**, 205–215 (2001).

11. W.W. Cooper, R.G. Thompson, R.M. Thrall. Introduction: extensions and new developments in DEA. *Annals of Operations Research* **66**, 3–45 (1996).

12. R. Färe, S. Grosskopf. Slacks and congestion: a comment. *Socio-Economic Planning Sciences* **34**, 27–33 (2000).

13. R. Färe, S. Grosskopf. Research note: decomposing technical efficiency with care. *Management Science* **46**, 167−168 (2000).

14. R. Färe, S. Grosskopf, C.A.K. Lovell. *The Measurement of Efficiency of Production*. Kluwer-Nijhoff, Boston (1985).

15. R. Färe, S. Grosskopf, C.A.K. Lovell, C. Pasurka. Multilateral productivity comparisons when some outputs are undesirable: a non-parametric approach. *Review of Economics and Statistics* **71**, 90–98 (1989).

16. A.T. Flegg, D.O. Allen. Does it matter how we measure congestion? Discussion paper 06/14. School of Economics, Bristol Business School, University of the West of England, Bristol (2006). Available online at: http://carecon.org.uk/DPs/.

17. A.T. Flegg, D.O. Allen. Does expansion cause congestion? The case of the older British universities, 1994−2004. *Education Economics* **15**, 75−102 (forthcoming March 2007).

18. A.T. Flegg, D.O. Allen, K. Field, T.W. Thurlow. Measuring the efficiency of British universities: a multi-period data envelopment analysis. *Education Economics* **12**, 231−249 (2004).

19. S.C. Ray. *Data Envelopment Analysis: Theory and Techniques for Economics and Operations Research*. Cambridge University Press, Cambridge (2004).

20. K. Tone, B.K. Sahoo. Degree of scale economies and congestion: a unified DEA approach. *European Journal of Operational Research* **158**, 755–772 (2004).

21. J. Zhu. *Quantitative Models for Performance Evaluation and Benchmarking: Data Envelopment Analysis with Spreadsheets*. Kluwer, Boston (2002).

Table 1. Results from Cooper's approach: textile industry

	CCR efficiency score	BCC efficiency score	Scale efficiency score	Labour congestion	Capital congestion	Overall congestion
1981	0.276	1	0.276	0	0	0
1982	0.263	0.669	0.393	0	0.034	0.017
1983	0.288	1	0.288	0	0	0
1984	0.325	1	0.325	0	0	0
1985	0.493	1	0.493	0	0	0
1986	0.447	0.747	0.598	0	0	0
1987	0.433	0.552	0.785	0	0	0
1988	0.356	0.361	0.985	0.091	0	0.046
1989	0.396	0.426	0.930	0.061	0.020	0.041
1990	0.613	0.654	0.938	0.058	0	0.029
1991	0.859	1	0.859	0	0	0
1992	0.809	0.874	0.925	0.041	0	0.021
1993	0.688	0.716	0.961	**0**	0	**0**
1994	0.956	1	0.956	0	0	0
1995	0.857	0.937	0.914	0	0.136	0.068
1996	1	1	1	0	0	0
1997	1	1	1	0	0	0

Table 2. Results from Cooper's approach: automobile industry

	CCR efficiency score	BCC efficiency score	Scale efficiency score	Labour congestion	Capital congestion	Overall congestion
1981	0.873	1	0.873	0	0	0
1982	0.938	1	0.938	0	0	0
1983	1	1	1	0	0	0
1984	0.961	0.999	0.962	0.121	0	0.061
1985	0.818	0.851	0.961	0.203	0	0.102
1986	0.658	0.674	0.976	0.133	0	0.067
1987	0.603	0.614	0.982	0.135	0	0.068
1988	0.690	0.706	0.978	0.194	0	0.097
1989	0.874	0.902	0.969	0.237	0	0.119
1990	0.830	0.853	0.973	**0.219**	0	**0.110**
1991	0.857	0.876	0.978	0.239	0	0.119
1992	0.861	0.871	0.989	0.187	0	0.094
1993	0.828	0.830	0.987	0.073	0	0.036
1994	0.839	0.839	1.000	0.008	0	0.004
1995	0.961	0.971	0.990	0	0.139	0.069
1996	0.940	0.940	1.000	**0**	0	**0**
1997	1	1	1	0	0	0

Table 3. Results from Färe's approach: textile industry

	Overall technical efficiency	Pure technical efficiency	VRS-based method		CRS-based method	
			Scale efficiency	Congestion efficiency	Scale efficiency	Congestion efficiency
1981	0.276	1	0.276	1	0.969	0.284 K
1982	0.263	1	0.279	0.945 K	1	0.263 K
1983	0.288	1	0.288	1	0.288	1
1984	0.325	1	0.325	1	0.365	0.890 K
1985	0.493	1	0.493	1	0.548	0.900 L
1986	0.447	0.944	0.473	1	0.524	0.903 L
1987	0.433	0.864	0.501	1	0.540	0.928 L
1988	0.356	0.707	0.503	1	0.510	0.985 L
1989	0.396	0.694	0.571	1	0.574	0.995 L
1990	0.613	0.837	0.733	1	0.793	0.925 L
1991	0.859	1	0.859	1	1	0.859 L
1992	0.809	0.935	0.865	1	0.945	0.915 L
1993	0.688	0.836	0.824	1	0.824	1
1994	0.956	1	0.956	1	0.956	1
1995	0.857	1	0.981	0.873 K	0.983	0.871 K
1996	1	1	1	1	1	1
1997	1	1	1	1	1	1

Table 4. Results from Färe's approach: automobile industry

	Overall technical efficiency	Pure technical efficiency	VRS-based method		CRS-based method	
			Scale efficiency	Congestion efficiency	Scale efficiency	Congestion efficiency
1981	0.873	1	0.873	1	1	0.873 L
1982	0.938	1	0.938	1	1	0.938 L
1983	1	1	1	1	1	1
1984	0.961	1	0.962	0.999 L	0.961	1
1985	0.818	0.826	0.990	1.000 L	0.990	1
1986	0.658	0.747	0.882	0.999 L	0.881	1
1987	0.603	0.736	0.823	0.996 L	0.820	1
1988	0.690	0.710	0.972	1	0.972	1
1989	0.874	1	0.979	0.893 L	0.874	1
1990	0.830	0.844	0.988	0.995 L	0.983	1
1991	0.857	0.918	0.986	0.946 L	0.933	1
1992	0.861	0.883	0.994	0.982 L	0.976	1
1993	0.828	0.836	0.990	0.999 K	0.990	1
1994	0.839	0.909	0.929	0.999 K	0.923	1
1995	0.961	1	0.976	0.984 K	1	0.961 K
1996	0.940	0.959	0.982	0.998 K	0.980	1
1997	1	1	1	1	1	1

Table 5. Results from Tone's approach

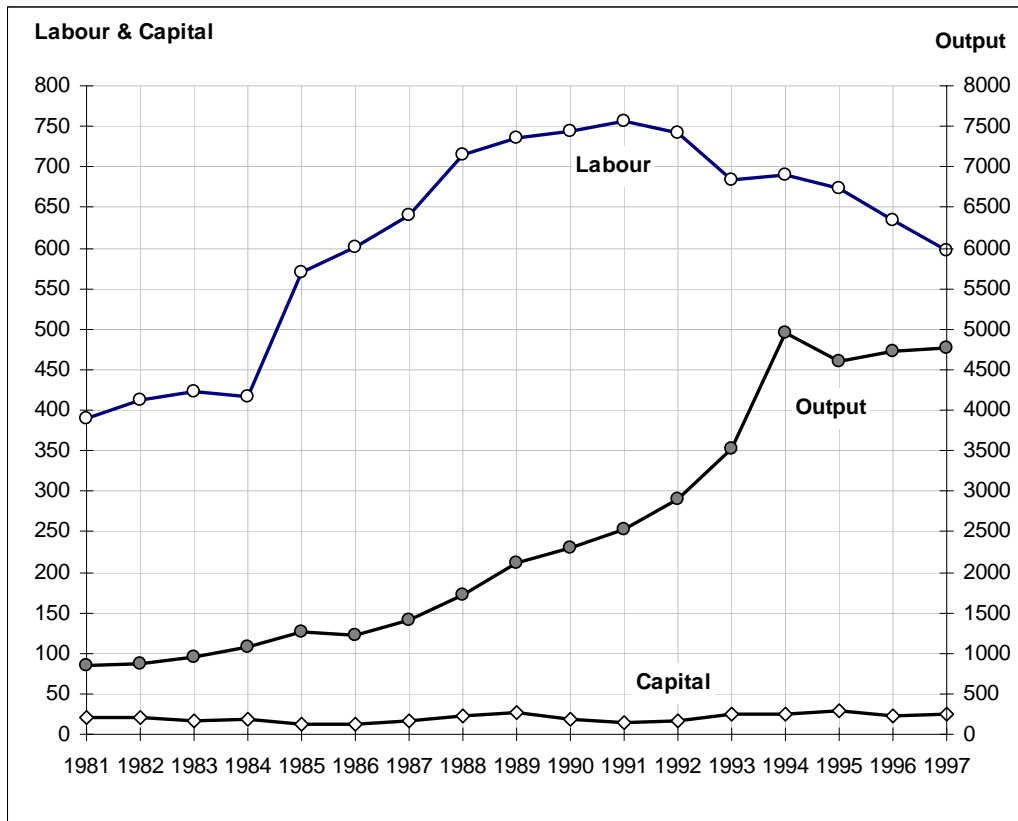|	Textiles		Automobiles	
	Congestion score	Scale diseconomy	Congestion score	Scale diseconomy
1981	1		1	
1982	1.494	−14.55	1	
1983	1		1	
1984	1		1.001	−0.01
1985	1		1.175	−0.86
1986	1		1.023	−0.18
1987	1		1.025	−0.19
1988	1.291	−3.19	1.066	−0.34
1989	2.346	−33.18	1.109	−0.46
1990	1.315	−5.43	1.091	−0.41
1991	1		1.141	−0.59
1992	1.144	−3.49	1.148	−0.79
1993	1.002	−0.58	1.204	−2.81
1994	1		1.002	−0.22
1995	1.067	−0.50	1.030	−0.22
1996	1		1.001	−0.09
1997	1		1	
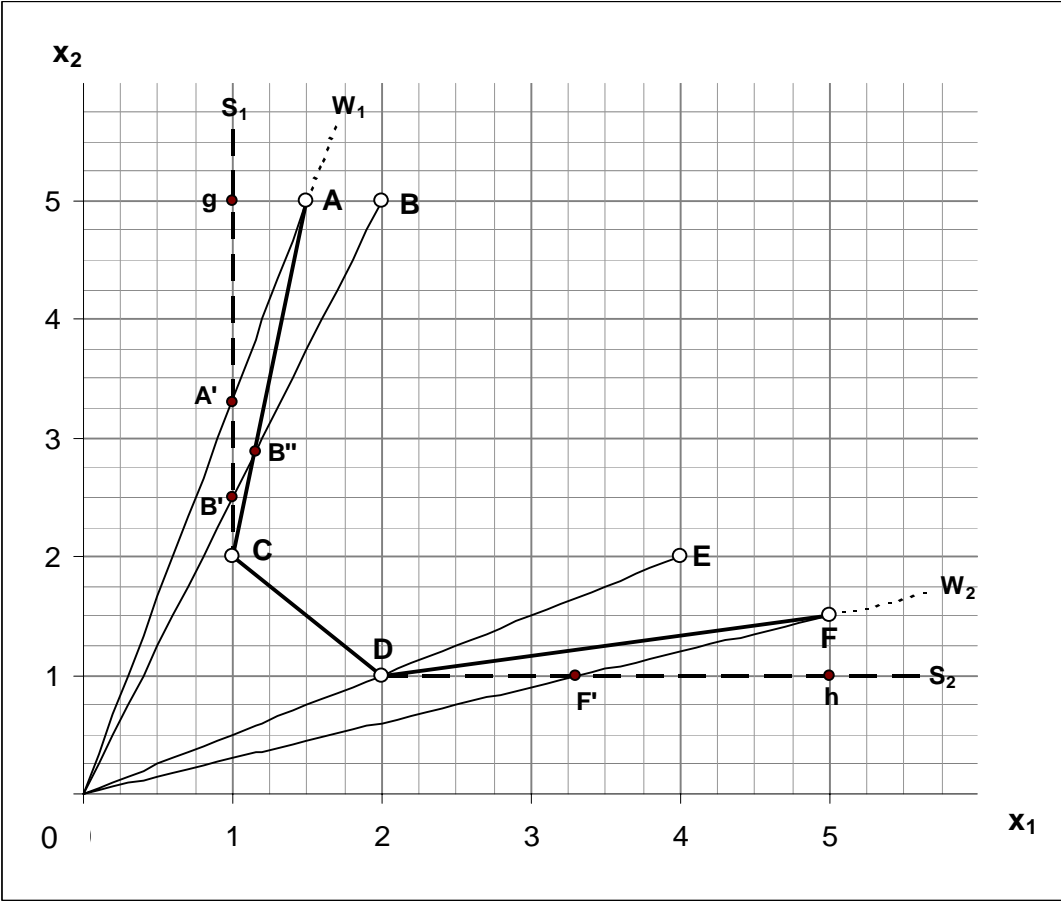
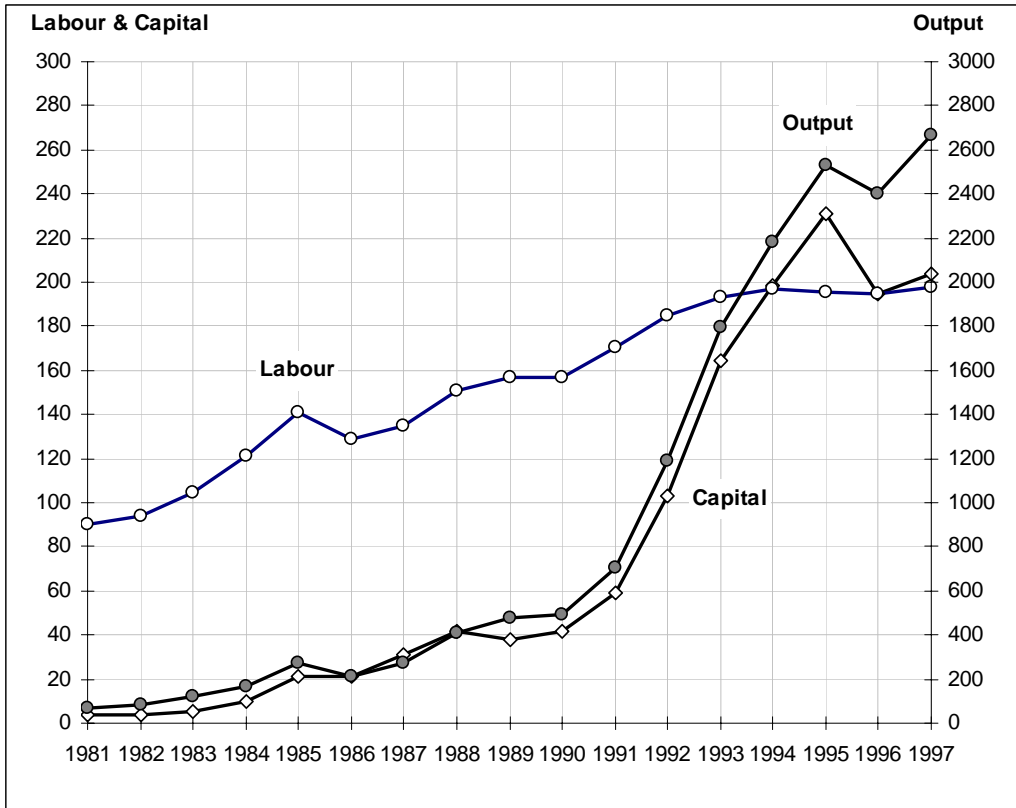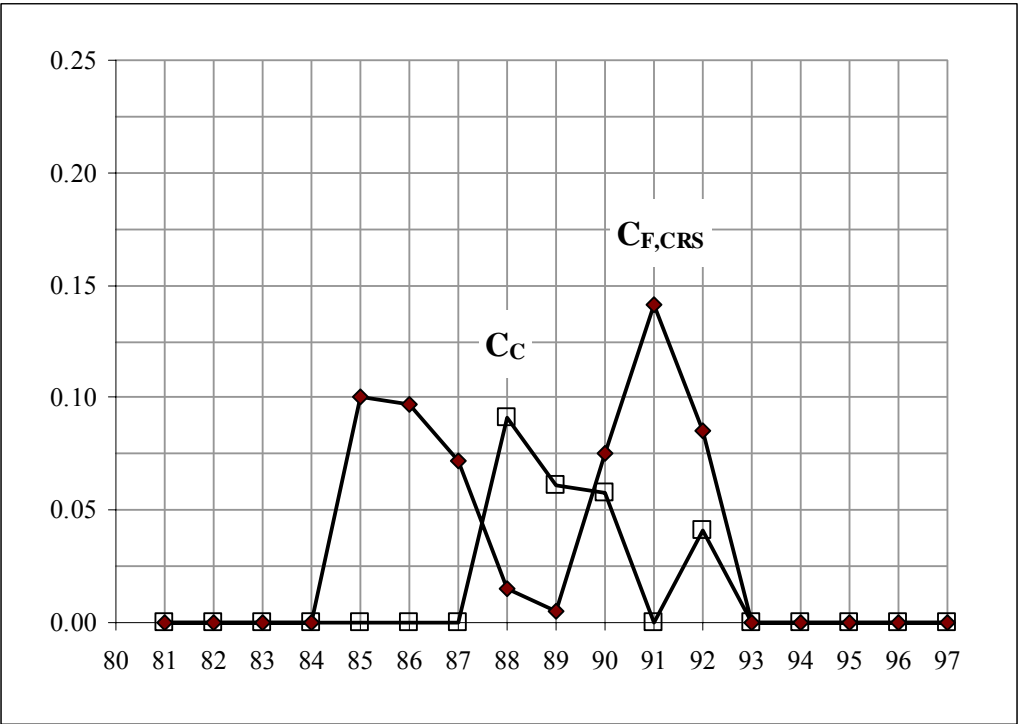**Figure 1.** Färe and Grosskopf's approach
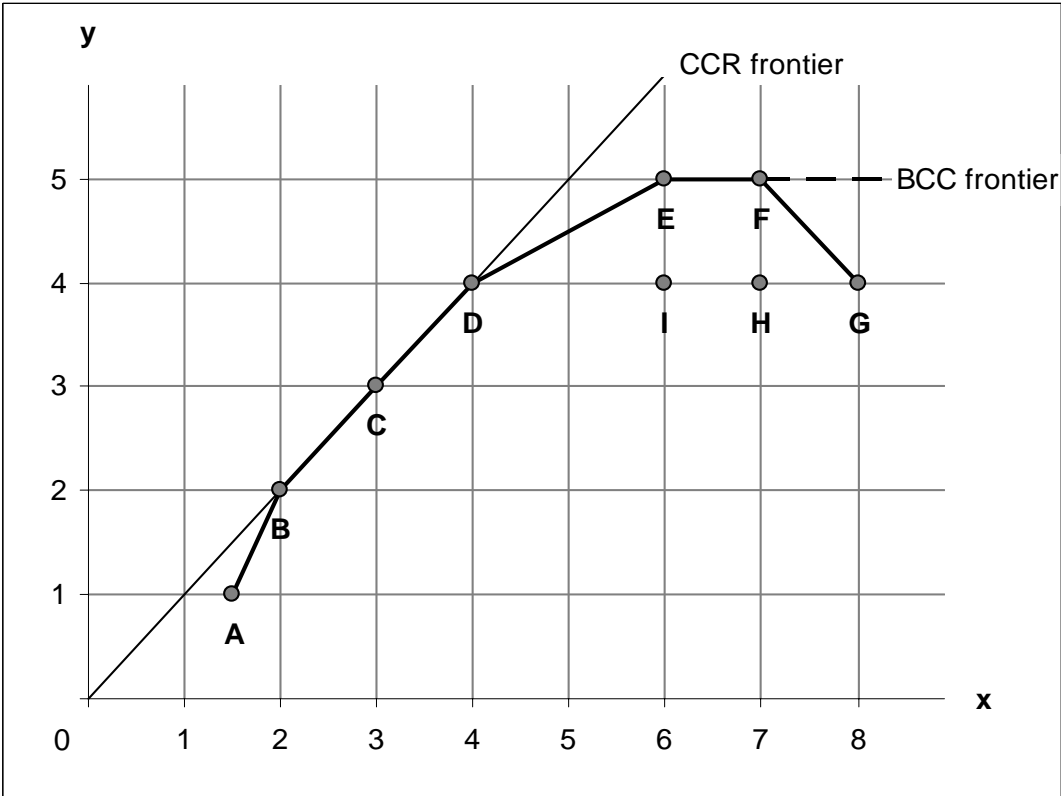
Figure 2. Cooper's model
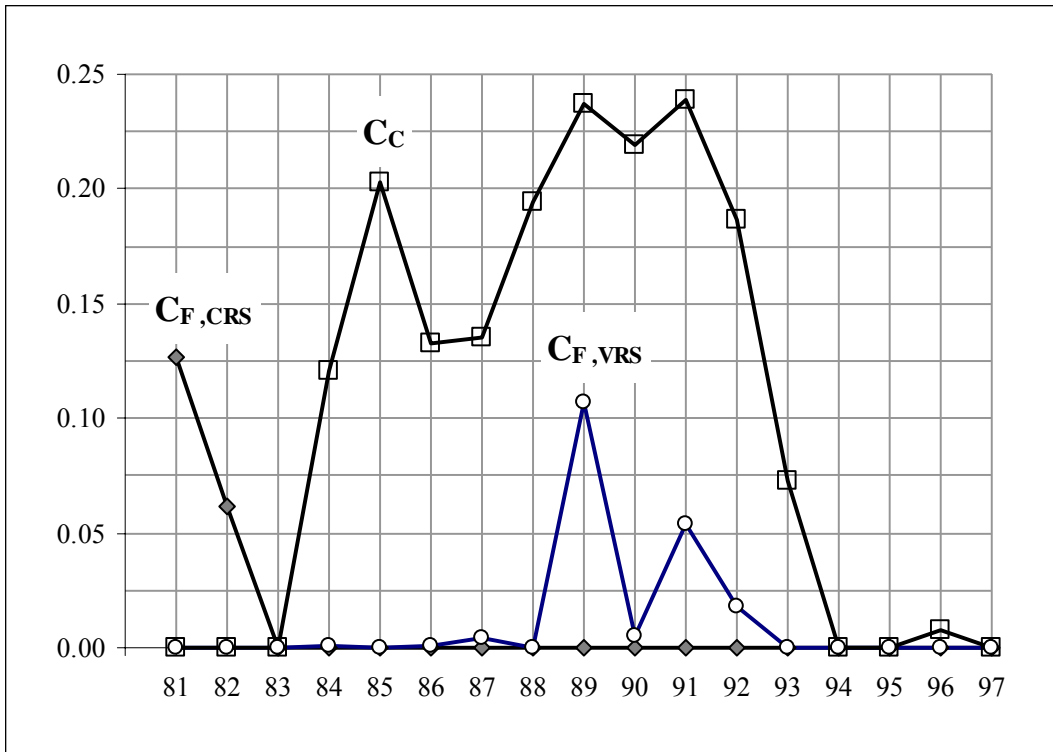
Figure 3. Data for the textile industry

Figure 4. Data for the automobile industry

Figure 5. Alternative measures of labour congestion: textiles

Figure 6. Alternative measures of labour congestion: automobiles