Rejoinders to the comments on my paper "Performance measurement and joint production of intended and unintended outputs"

Finn R. Førsund[1]

Abstract

I will comment on the comments by the groups of three reviewers separately. First of all, I will thank all three groups providing a first round of reports in order for me to get rid of obvious mistakes. In the second round the reviewers were free to comment on the qualities of my revised version. I am not to change my revised paper when giving my comments on what would be honest reports on the quality of my final version. However, the reviewers will not have a go at my rejoinders to comments according to the symposium rules. I keep the section numbering of the authors in order to make it easier to identify the arguments.

The comments of Rolf Färe and Shawna Grosskopf

Acknowledging that I and Murty and Russell (2002), Murty, Russell and Levkoff (2012), followed by several other papers, have developed a model with two separate production functions, Rolf and Shawna leave the discussion of separate production functions to Murty and Russell. Instead they focus on the discussion of the 'Shephard inspired approach' and its development since the 1980's with focus on specification and linkage of production and abatement nodes in a network, and how the network model can be regarded equivalent to multi equation models. They provide a unifying framework for modelling goods and bads by specifying multiple sub technologies which can be linked and estimated with a distance function.

1 Production functions and subtechnologies

Concerning the single equation model inspired by Shephard (1970) I do not insist that the Frisch inspired factorially determined multi output production function is necessary, but it is sufficient to capture joint production of the type that production of bads is *unavoidable*. The two types of production functions in Equation (9) in my paper do not represent two independent functions; x_M must be the same in both equations. To split into two functions a process that happens simultaneously is done in order to state that the amount of the good output is independent of the amount of the bad output and vice versa. Process engineers will probably not find it interesting in an engineering sense, for them the importance is to design a process that maximises the amount of good output and minimises the bad output for given non-material and material inputs. I regard the $f(.)$ and $g(.)$ functions as frontier functions where the good y is the maximal amount produced by $f(x_M, x_S)$ and the bad z is the *minimal* amount produced by $g(x_M)$ when you have the condition that z is greater or equal to the frontier value given x_M. I see no problem here, cf. the shape of cost functions.

Looking at Fig. 1 in Rolf and Shawna comment it seems that inputs are distributed on subtechnologies P^1

✉ Finn R. Førsund
finn.forsund@econ.uio.no

[1]   Department of Economics, University of Oslo, Oslo, Norway

and P^2. The authors say that "P^1 could represent the subtechnology producing good output, and P^2 the subtechnology producing unintended bad output". However, this is not compatible with what I define as joint production. In my understanding of unavoidable joint production of bads we must have $x_0 = x_0^1 = x_0^2$ (valid for x_{M0}); by definition of technical jointness (Frisch 1965) you cannot produce the good outputs without simultaneously producing the bads.

2 Weak disposability

The authors quote from my paper where I state that I am only interested in "residuals jointly with good outputs, where the residuals cause environmental damage, identified as positive willingness to pay to reduce the damage". They seem to claim by an illustration in their Fig. 2 that in that case no bads will be produced; bads are disposable. However, my point, made already in Førsund (2009), is that if you have *assorted* production (Frisch 1965), i.e. inputs can be reallocated to different outputs also including bads, then it will not be optimal to produce any bads. My point is that you must have the case of *technically joint production* (Frisch 1965) in order to capture the situation with unavoidable bads, or as introduced in Murty et al. (2012), *costly disposability*.

3 Null-jointness and material balance

The authors are introducing null-jointness of the good and the bad. This is not wrong for most cases, however, in my view the basic information is the connection between the *material inputs* and good and bad outputs. If nothing can be produced without material inputs, both goods and bads will be zero. The relationships between goods and bads go through material inputs. However, it may be the case that the bad output is zero even when the use of material input is positive. If a carpenter sources the right dimensions of planks from a sawmill he can make a very rough wooden table without creating any residuals within his workshop; positive material input is essential for positive good output, but no bad is created.

Concerning the limit of the bad in my paper this is based on the limit of the use of material inputs. I use these basic input limits in the bads function $g(.)$ in my modelling.

4 Network models: providing common ground?

Rolf and Shawna pioneered networks models starting with Färe and Grosskopf (1996). The focus was on independent firms selling intermediate outputs to firms producing the final outputs. Efficiency scores could then be calculated for each plant in the chain. Now the focus is on subtechnologies

within the same plant. They suggest that by creating an internal network model for subtechnologies can have a common ground with using two separate model types. I think this is a good idea, because subtechnology P^1 in Fig. 1 can represent the production of the good and the gross bad using two production functions as in Eq. (9) in my paper. Subtechnology P^2 representing abatement (treatment) can be modelled by adding an abatement function discharging the net bad (see e.g. Førsund (2009) where an abatement function with the net pollutant as the product and the gross output as an input together with relevant material and non-material and inputs).

References not used in the comments by Färe and Grosskopf

Färe R and Grosskopf S (1996). Productivity and intermediate products: a frontier approach. *Economics Letters* 50 (1), 65–70

Frisch R (1965). *Theory of production*. D. Reidel, Dordrecht

Førsund FR (2009). Good modelling of bad outputs: pollution and multiple-output production. *International Review of Environmental and Resource Economics* 3(1), 1–38

The comments of Frederic Ang and K. Hervé Dakpo

In their introduction Ang and Dakpo make a brief, but correct, summation of my paper arguing for a two-relation model. I appreciate that they state my emphasis on the importance of assorted production not being able to represent joint production. The domination of Shephard-inspired models based on weak disposability and null-jointness of the good and the bad is acknowledged. A separation of good and bad production has also been proposed by Murty and Russell in several papers. Such a model is called the by-production model.

1 The materials balance principle as an "accounting identity"

In my paper I emphasis that the mass balance equation is an identity. Ang and Dakpo seem to have another opinion. I am afraid I cannot follow them here. Several authors, among them Pethig (2006); Ebert and Welsch (2007) use the materials balance as part of making what are their production functions. However, this is not a good idea because the material balance is an *identity* expressed by a *linear* summing-up relationship of types of material substances and not a production function.

In Footnote 8 in my paper I express the problems arising not acknowledging that the mass balance equation is an identity in Coelli et al. (2007):

In Coelli et al (2007) using the materials balance, it is stated that "the only efficiency score that is consistent with the materials balance condition is a value of one, implying that inefficient production is not permitted" (p. 6). This focus is unfortunate because the materials balance also applies to inefficient points; the mass of input materials of inefficient points is also distributed on outputs and residuals by a linear identity. However, the task in Coelli et al (2007) is to minimise the content of a pollutant given output, and the linearity of the materials balance is then maintained.

However, the rest of the comments by Ang and Dakpo are devoted to themes not addressed in my paper. They argue that the materials balance equation reflects a duality between the material balance-based approach of Coelli et al. (2007) and the bad output technology of Førsund, and Murty et al. (2012). I have problems following them on this duality.

References not in the Comments by Ang and Dakpo

Ebert U and Welsch H (2007). Environmental emissions and production economics: implications of the materials balance. *American Journal of Agriculture Economics* 89 (2):287–293

Pethig R (2006). Non-linear production, abatement, pollution and materials balance reconsidered. *Journal of Environmental Economics and Management* 51(2), 185–204.

The comments of Sushama Murty and R. Robert Russell

There is an expression in German for the type of critique Murty and Russell give my paper: *"eine grausame salbe"*. I will try to explain some of the weaknesses of the paper in the light of the critique, stating "many errors, omissions, and misunderstandings". I will try to acknowledge my misunderstanding and errors in the paper where these are apparent to me. I go section by section and start with Murty and Russell statements in Italic that I want to react to.

1 Overview

We are of the opinion that almost all of the issues discussed in this paper are not new: I was invited to give a keynote lecture at the European workshop on efficiency and productivity analysis (EWEPA) in London 2019. I took the opportunity to brushing up old issues in my papers on "goods and bads" and also to try to give some new insights. Given such an opportunity at a conference it is not uncommon to use issues explored in previous own papers.

Although issues may be old ones, my intention was that the paper should improve upon issues making them more accessible than in my earlier papers. I put a special emphasis in Subsection 4.2 on the meaning of joint production where the bad outputs are *unavoidable* that I think have some novel information. Fig. 3 and the explanation of why an output isoquant (trade-off curve between intended and unintended outputs) is not possible in the case of technical jointness are quite new insights. However, my main focus in the present paper is to make a critique of the Shephard-inspired approach used by most authors following the seminal contributions of Rolf and Shawna and associates based on Shephard (1970). It has taken a long time accepting critique of the single equation model, but the by-product model of Murty et al. (2012) is now beginning to be used, and my approach is also included in some recent papers (Ang and Dakpo (2019); Aparicio et al., 2020).

In my critique of the Shephard approach I also make frequent citations of Murty and Russell and coauthors that have made powerful and mathematically elegant analyses of an approach they named *by-production*. By a formal mathematical analysis Murty et al. (2012) proved that this requires two separate production functions for goods and bads, respectively, in contrast to single equation models popular with authors following the Shephard approach. I used another approach to realise that separate production functions were required based on *factorially determined multi-output production functions* introduced by Frisch (1965); one of the giants of economics, that still gave lectures now and then when I was an economics student in Oslo.

… the approach for modeling emission-generating technologies advocated by the author is converging absolutely to the most primitive case of the by-production approach first proposed by Murty and Russell [2002], empirically implemented by Murty, Russell, and Levkoff [2012]: Starting to expand the Frisch multi equation model in Førsund (1998) my revised model in Subsection 4.2 is called "the most primitive case of the by-production approach". However, I just follow Murty et al. (2012):

"In order to strip the argument to its barest essentials, we consider a very parsimonious model in which two inputs—one pollution generating, the other not—are employed to produce a single intended output and a single unintended output (p. 119)".

In my Subsection 4.2 I state: "In order to illustrate the model [in Eq. (9)] in the simplest way I specify only two

outputs, y and z, and two inputs, x_M and x_S", and I continue in a parenthesis: "the number of intended and unintended outputs and factors of production can easily be extended as shown above in Eq. (8a, b)". The last statement is also mentioned in my Conclusions in Section 7.

The difference between the two models is explained in my paper in Subsection 4.3 on the by-production model:

"The by-production model is quite close to the factorially determined multi-output model in Subsection 4.2 regarding the splitting into two types of production functions. However, the two model types are based on different arguments. The factorially determined multi-output model was introduced by Frisch (1965) as a variant of his general multi-output model (7) for technically connected outputs. Production functions are specified for each output, thus separating outputs. Technical jointness is maintained by having the same bundle of inputs in all functions. Only intended outputs were considered, but an extension to include unintended outputs as done in Subsection 4.2 seems obvious (but need a modification as to the role of service inputs as seen in Eq. (9))".

2 Material balance

… *in practice it might be impossible for the researcher to spell out completely the material-balance conditions*: Murty and Russell interpret my statements on the difficulties of empirically represent the materials balance at the end of Section 2 in a way making excuses for features of the Shephard approach I criticise. I copy my statements below:

… "measuring all the factors involved in the materials balance accurately may not be so easy, especially on the more aggregated level that is commonly used in efficiency analyses. If we accept that residuals are measured accurately, we know that all observations of production units, efficient as well as inefficient units, must obey the materials balance as an identity. If we do not have observations, but data that are theoretical it may not be feasible to assign the materials balance accurately to hypothetical observations based on observed ones".

My critique is directed to theoretical models. A typical problem is revealed in my Fig. 1, copied from Shephard (1970), concerning the trade- off curve between a good output and a bad one. I take this figure as representing a *theoretical* model. Then I think it is evident that the materials balance is not fulfilled because the input is constant along the trade-off curves by definition, while both the good and bad output decreases towards the origin. I cannot see that the empirical problem of accurately observing the materials balance is relevant for this violation of the material balance illustrating a theoretical model.

3 Frisch models of multi-output technologies

… *all the cases studied by Frisch and reported in the current and previous works of the author are nothing more than suitable applications of the implicit function theorem:* I think the way Murty and Russell use the implicit theorem to interpret Frisch gives a new and very useful characterisation. I have not seen this before in any publication. Using the implicit function theorem on the Frisch multi-equation system $F^k(x, y) = 0$, $k = 1, …, \mu$, Murty and Russell find both the case of jointness and maximal assortment.

4 Weak disposability

The author's earlier exposition of the Shephard [1970] weak-disposability model in Section 3.1 is confusing (Footnote 4): it may be confusing, but unfortunately I fail to see how this can be misunderstood, my purpose is to expose what I think is problematic with the Shephard approach, so I may not have a sufficiently rigorous exposition as experts have.

… *boundary of the sets represents efficient operations* (Footnote 4): I do not follow a strict axiomatic building-up, I assume that the typical reader understands the role of the boundary. Shephard (1970, Chapter 4.4) states in the interpretation of Fig. 1 that points on the solidly drawn boundary are efficient.

… *the points on the vertical segments of the boundaries in Figure 1, contrary to the author's declaration, are not efficient* (Footnote 4): I cannot see that I have said that the vertical segments in the Shephard Fig. 1 are efficient, as far as I can see these lines are thin and I have interpreted that Shephard assume that thin lines are not efficient. But according to a communication from Rolf and Shawna in my footnote 12 the thin lines emanating from the origin are efficient.

… *as abatement activities draw away the resources of the production unit to mitigate emissions, lesser quantities of both the good and the bad outputs are produced*: In my paper I take isoquants at face value. To introduce abatement that is based on reallocating resources from production of goods and bads should not interfere with the shapes of isoquants in output space; allocating inputs to abatement can be understood as doing this before drawing isoquants. It seems very awkward to do a simultaneous change in abatement and outputs, at least it can hardly be drawn. As I have stated in my paper the materials balance is violated using Shephard type of trade-off curves in output space. By *definition* an output isoquant has a constant amount of inputs (isoquant means *same quantity*; Frisch introduced

this term). I cannot see that the problem introducing abatement has been solved. It does not help to "implicitly assume[d] that abatement activities are increasing", as stated by Murty and Russell. It is simply not possible to avoid the three points Murty and Russell state taken from my paper.

5 Measurement of efficiency and productivity

Section 6.2 on the measurement of efficiency and productivity is flawed and misleading; the author would be well advised to replace this subsection with references to the salient literature: Murty and Russell stated in the first round of a review report that I should drop this subsection and refer to the literature instead. I have therefore shortened the subsection considerably referring to DEA literature for setting up the appropriate LP programs, and just mention that $E_y(y^*/y^{obs})$ and $E_z(z^{obs}/z^*)$ are the Farrell radial efficiency scores and that the Malmquist productivity indices are the change in the ratio of these efficiency indices from one period to the next. I have difficulties in believing that a reader can misunderstand this. I found it necessary to just formulate these basic things because productivity measures build on efficiency measures, and the productivity measure for a bad has not been so common. It should be quite obvious that I, with many journal publications on efficiency and productivity, consciously avoided stating strictly formal definitions of efficiency and productivity measures that Murty and Russell now want.

Note that these index-number formulations do not require the author's assumption of constant returns to scale: Grifell-Tatjé and Lovell (1995) point out that specifying variable returns to scale the Malmquist productivity index does not accurately measure productivity change. Therefore, I stick to constant returns to scale.

6 By-production

… without attribution: I have referred to Murty and Russell many times in my paper, and I can only apologise if I have missed to refer sometimes, I assure the authors that this was not intentional.

No, the MRL [2012] critique not only corresponds to, but also substantially pre-dates the author's discussion in Subsection 4.2 (Footnote 8): As said above, I apologise for not referring to MRL (2012), but after all, I also criticise Baumol and Oates (1988) in Førsund (2009) and not for the first time in Subsection 4.2. I have lectured an environmental economics course for many years, and used the Baumol and Oates (1988) book specially to tell the students about the mistakes in the externalities chapter about goods

and bads, maybe my experience made me to forget what you had done about Baumol and Oates (1988) in your (2012) paper. "Corresponds" does not mean that this take place before Murty and Russell (2002); MRL (2012), it is just the same story.

The author of this paper defines and refers to these two upper and lower frontiers of the technology several times (see e.g., Section 4.2) without citing this earlier work: In Murty et al. (2012) the upper and lower frontiers are mentioned in Footnote 21, p.125, and the figures shown in Fig.1 on the next page. However, this follows rather straightforwardly from the production possibilities sets for the two types of outputs, so I feel that it is not such a sin to have forgotten to refer to Murty and Russell here.

Førsund's approach to this issue adding the constraint, $\bar{z} \geq z \geq g(x_M)$, is at best problematic and misleading, since $\bar{z}$, as well as z, depends on x_M, the levels of the material inputs: I agree with the statement. I meant that $\bar{z}$ is a function of x_M, as expressed by the text immediately below the Eq. (11):

"Here $\bar{z}_i$ is the total material contained in x_{Mi}. If we consider only one type of substance for convenience, we have from the materials balance Eq. (1) that $\bar{z}_i = ax_{Mi}$. Obviously, the maximal amount of residual cannot be greater than this amount, but will be less if the intended output contains materials".

This statement is false: I used Russell (1998) as my source for difficulties of multiple equations in footnote 15:

"However, if a single transformation function is used with several outputs as arguments, maximising one intended output at a time keeping the other intended outputs constant for given inputs results in a different production function for each intended output, complicating the usefulness of a single transformation relation. See Russell (1998) for difficulties expressing joint production functions with many outputs".

Unfortunately, I am not much of a mathematician. I am convinced by the statement of Murty and Russell, I can only say that I have not seen the following statement before:

"the frontiers defined by these production functions contain the same information: any one of these production functions suffices to obtain a representation of the technology set. A standard neoclassical technology with a single output or multiple outputs requires only one production function for its representation".

Førsund [1998/2016, 2009, 2018] specifies two production relations of the form $y = f(x_M, x_S)$ and $z = g(x_M, x_S)$: These papers are written before I realised that it was not

correct to use x_S as an active input in the bads function thanks to Russell pointing this out as my discussant at the EWEPA 2019 workshop in London. I am surprised that Murty and Russell also comment at length on my previous papers having this flaw. After all, it is the present paper that is to be criticised.

This is done without explanation and without acknowledgement that this revision brings his model into conformity with the specification of Murty and Russell [2002] and MRL [2012]: The explanation is done in a sentence just before Eq. (9):

"However, the functions $g^k(.)$ is only influenced by the material input x_M due to the materials balance; the materials in z can only come from the material inputs".

Bringing into conformity is done in Subsection 4.3 by the statement:

"The by-production model is quite close to the factorially determined multi-output model in Subsection 4.2 regarding the splitting into two types of production functions".

However, I can understand if "quite close" is too weak a characterisation for Murty and Russell. I was thinking about different ways of introducing multiple good and bad outputs of the two models.

Acknowledgement (Footnote 13): I was under pressure from reviewers to shorten my paper. I shortened somewhat the footnote acknowledging Russell's role in giving me the correct understanding, and placed it in another page in the first paragraph after Eq. (10) than the placement in the first version submitted to JPA. The new text reads:

"The unintended output z is generated only by the materials input. If we can call it an isoquant in a two-input case it has to be vertical, as will be shown in Fig. 2".

My acknowledgement of Russell now reads in my footnote 31 to the quote above:

"Unfortunately, this insight, due to comments made by Robert Russell at the EWEPA meeting in London 2019, is not present in Førsund (2009); (2018a, b, c)".

I hope Russell accepts this acknowledgement of him that he seems to have overlooked.

There is, however, a fundamental misconception in Sections 4.2 and 4.3 of his paper: I agree with the explanation of Murty and Russell of my misconception, I wish I had the same level of understanding.[1]

… we cannot make sense of Figure 2. The main problem is that, while the y-isoquants are stationary in this space,

the so-called z-isoquant would have to shift as the input bundle (x_M, x_S) moves along either of the isoquants: Yes, the shift of the vertical isoquant is correct, but I cannot see that this should be a problem. I have consciously used the formulation cited above: "If we can call it an isoquant" for the solid vertical line in Fig. 2, and I have tried to give explanations of how to understand the figure. Unfortunately, the explanation of moving the vertical "isoquant" of the $g(.)$ function does not say explicitly that the dotted vertical line up to the point I from the abscissa is the isoquant for the x_M value indicated on the abscissa. I have probably thought that it was self-evident that the vertical isoquants follow the value of x_M.

7 Abatement

In the environmental economics textbook literature, a standard modelling of abatement has been the end-of-pipe as a separate equation with primary (gross) pollutants as inputs together with standard, but maybe specific inputs for abatement, and secondary (net) pollutants as the output. This is the treatment function. Another option is to do prevention (internal abatement). The most common such effort seen in the literature is improvement of the technology (see e.g. Porter and van der Linde 1995). Using gas instead of coal, light oil instead of heavy oil, etc., is also prevention. However, it is not so common to call substitution between inputs in use as prevention. The mix of inputs is determined by cost if no regulation. In my paper I use the Frisch (1935) chocolate example of substitution between chocolate mass and labour as an example of ex post substitution. Internal recirculation of raw materials, e.g. wood fibres when making paper, is a characteristic of the technology.

Such substitution between material and service inputs shown by the good output isoquants in my Fig. 2 is also prevention, but is not in the category of changing technology.

I must admit that although I have read both Murty and Russell (2018); (2017)[2] I could not see that the mathematics also covered both preservation and treatment. I must also admit that I am a little confused about the meaning of y^a and a. Murty et al. (2012) state (p. 121): "We model abatement activities as an output, y^a, that is used to mitigate pollution". However, in the frontier function $z = g(x_2, y^a)$ the output y^a functions as an *input*. In Murty and Russell (2018) y^a is not used, a is now used, and called cleaning-up output and also abatement output. Finally, Murty and Russell (2017), Section 10.5.1, give a clarification that I think is correct:

---

[1] I have had to change the numbering of a few equations in the proofs of my main paper, so the Murty et al. (2012) model referred to in Murty and Russell's remark as Eq. (12) is now Eq. (13).

[2] I refer to an early version dated May 2017 of the book Chapter 10 of "Bad Outputs". I see in the reference of Murty and Russell that the chapter is now 12, and that a first online version should be out in 2020. Unfortunately, I have not been able to get this on internet. I assume that the final version is not too different.

"The abatement technology employs inputs such as labour, capital, and lime or limestone to produce reductions in the emission. … Thus, the abatement technology is defined by relations among all the inputs used by it and the extent of reduction that is made possible by usage of these inputs".

However, as stated above, I prefer to model the treatment of emission as end-of-pipe that is discharged to the environment; this secondary (net) pollutant is a function of primary (gross) pollution and inputs needed for transforming a substantial part of the primary pollutant to un-harmful emissions (at least less harmful than the primary pollutant). Prevention is of another nature as stated above.

This is false on both counts: The explanations that Murty and Russell give concerning prevention and treatment are most enlightening. However, I think there may be readers of my paper that also have problems seeing prevention and treatment in the papers of Murty et al. (2012); Murty and Russell (2018); (2017).

… in the empirical section of their by-production paper, MRL [2012] model substitution among multiple fuels, coal, oil and natural gas: Annual data for 92 coal-fired electric power plants from 1985 to 1995 are used. Electricity is the intended output and coal, oil, and natural gas are inputs. However, coal constituted at least 95% of the total fuel consumption. So there is some limited substitution between the three energy inputs that may be called prevention. It is the same situation as in Fig. 2 in my paper with substitution between material and service inputs. As stated previously this is not the type of prevention that is in focus in the environmental economics literature.

References not in the comments of Murty and Russell

Frisch R (1935) The principle of substitution. An example of its application in the chocolate industry. *Nordisk Tidskrift for Teknisk Økonomi* 1:12–27

Grifell-Tatjé E, Lovell CAK (1995) A note on the Malmquist productivity index. *Econ. Lett.* 47(2):169–175

Porter ME, van der Linde C (1995) Toward a new conception of the environment-competitiveness relationship. *J. Econ. Perspect.* 9 (4):97–118

Compliance with ethical standards