The Effects of R&D on Agricultural Productivity of Australian Broadacre Agriculture: A Semiparametric Smooth Coefficient Approach

Farid Khan

School of Economics and Finance, Curtin University, Western Australia

Email: faridecoru@yahoo.co.uk

&

Ruhul Salim

School of Economics and Finance, Curtin University,

Western Australia

Email: Ruhul.Salim@cbs.curtin.edu.au

Selected Paper prepared for presentation at the 2015 Agricultural & Applied Economics Association and Western Agricultural Economics Association Annual Meeting, San Francisco, CA, July 26-28

The Effects of R&D on Agricultural Productivity of Australian Broadacre Agriculture: A Semiparametric Smooth Coefficient Approach

Abstract: This article analyses the role of research and development (R&D) in Australia's broadacre farming by using the semi-parametric smooth coefficient model proposed by Hastie and Tibshirani (1993) and Li *et al* (2002) and a state-level dataset covering the period 1995 to 2007. While the conventional production function approach only captures the direct effects of R&D, this methodology captures both the direct impact of a change in R&D on output and the indirect impact through changes in efficiency of use of factor inputs in the production process. The empirical results show that once both the direct and indirect effects are taken into consideration, R&D investments significantly increase outputs. The results also show that there are substantial variations in the effects of R&D on output across the states. Such variations need to be taken into account when designing policies for investing public R&D in agriculture.

Keywords: Broadacre Agriculture, Semi-parametric smooth coefficient model, Productivity, Research and Development

JEL Classification: C14, C23, D24

1. Introduction

There is broad consensus among economists and researchers that the growth in agricultural productivity has been playing a leading role in meeting the growing global food demand (Alston and Pardey 2014; Fuglie and Toole 2014; Pardey *et al*. 2013). Over the past several decades, considerable research has been undertaken to analyse the impacts of research and development (hereafter, R&D) on total factor productivity (hereafter, TFP) in both the industrial and agricultural sectors. A number of studies provide empirical evidence that R&D is one of the primary sources of productivity growth (Alene 2010; Coe and Helpman 1995; Griliches 1998; Mullen *et al*. 2008). In agriculture, the role of public R&D in productivity has been recognized since the early studies of agricultural economics. For example, Schultz (1953) estimates the returns to public R&D and attributes all of the productivity growth in agriculture to public investments in agricultural research. Similarly, Griliches (1964) estimates the Cobb-Douglas type agricultural production function while introducing a research and extension variable along with the conventional input variables.

Recent studies have found a close correlation between investment in public R&D and TFP in agriculture. Studies such as Alston *et al*. (2011), Fuglie and Toole (2014) and Wang *et al*. (2013) provide evidence that R&D investments in agricultural research provide new knowledge and technologies that fuel improvements in agricultural productivity in US agriculture. Wang *et al*. (2013) has shown that R&D affects agricultural productivity only over the long-term. Changes in public R&D stocks have a significant impact on agricultural TFP growth. Similar evidence is also found for developing countries. For example, a study by Rahman and Salim (2013) in Bangladesh shows that R&D investment is one of the significant aspects that favourably affect TFP growth. Furthermore, Voutsinas and Tsamadies (2014) have found that R&D expenditure in Greek agriculture improves the rate of technological innovation, which affects long-run productivity growth.

Productivity growth in agriculture has been an essential source of economic prosperity in Australia. The contribution of R&D expenditure to farm productivity growth is also evident in Australian agriculture. According to studies by Mullen (2007, 2010), investments in agricultural R&D and policies that affect agricultural R&D are central to improvements in agricultural productivity growth in Australia. Investments in R&D lead to a more effective use of existing resources and thereby increase productivity levels. Using historical data and standard time series techniques, Salim and Islam (2010) find that R&D affects long-run productivity growth in agriculture in Western Australia.

In recent periods, there has been concern that the productivity growth in agriculture is slowing in developed countries. Australia has been facing slowing productivity growth in at least some sectors of agriculture (Islam *et al*. 2014; Khan *et al*. 2014; Sheng *et al*. 2014). Similar evidence of slowing productivity growth in recent periods has been seen in US agriculture (Ball *et al.* 2013). Piesse and Thirtle (2010) have also found a slowdown in TFP growth in agriculture in the United Kingdom. They mention a slowdown and retargeting of public R&D as one of the main factors causing this productivity slowdown. Similarly, other studies also suggest that one of the primary reasons for slowing productivity growth in agriculture is that public investment in R&D has been declining over the past few decades (Bervejillo *et al*. 2012; Mullen 2010; Pardey *et al*. 2013; Suphannachart and Warr 2011). These recent phenomena in agriculture have rekindled interest in investigating the relationship between public funding in agricultural R&D and productivity.

The conventional estimation of effects between R&D and productivity generally focuses around country-level or state-specific (i.e., for a particular state) data, but fail to reflect state-level technological heterogeneity. Farms face heterogeneous R&D environments across states, and R&D likely has differential effects on agriculture across different states in Australia. The agricultural structure, physical environment and market circumstances are different from one state to another, which has implications for productivity performance variations across states. Therefore, state-level variations need to be accounted for when estimating the impact of R&D on the productivity in Australian broadacre agriculture. This study aims to fill this empirical research gap in Australian agriculture.

In addition, while it is widely perceived that R&D makes significant contributions to agricultural productivity growth, research has rarely considered non-neutral effects of R&D in the empirical models of agricultural TFP growth. Studies capture the direct effect of R&D expenditure on productivity, but they fail to capture the indirect effects through the efficiency with which factor inputs are used. Therefore, the heterogeneous impact of the R&D on input productivity has largely been neglected in the previous empirical estimations. Furthermore, estimates of the effects of R&D on productivity that have been performed by researchers who apply parametric models are generally based on the assumption that the error term is normally distributed. The non-neutrality of technical change and the non-normality of errors in parametric models may prompt biased estimates of the R&D impacts because they depend on presumptions of the functional form and the distribution of the error term that cannot be known *a priori*.

Against theses backdrops, a number of studies have emerged in the broader economics literature that use semi-parametric or nonparametric approaches to address these problems (Mamuneas *et al.* 2006; Zhang *et al.* 2012; Zhao *et al.* 2014).

Parmeter *et al.* (2014) compared the parametric and nonparametric methods applying Norwegian dairy farm data and found that the nonparametric model provides improved out-of-sample prediction especially when employing the constrained estimator. The semi-parametric smooth coefficient model is one such empirical approach, and it has potential for the agricultural literature, particularly with regard to gaining a deeper understanding of the relation between R&D and TFP. The semi-parametric estimator used to estimate the marginal effects of R&D is the kernel density estimator, which avoids the functional forms and distributional assumptions of the parametric models and permits nonlinearity in the model. The main advantage of using this recent methodology is that it permits all sorts of nonlinearities and interactions between the factors without requiring any (preliminary) parametric formulation.

Unlike traditional inputs, such as capital, labour and materials, R&D is one of the environmental factors that characterize the production environment in general. A change in an environmental factor is likely to affect the productivity of the traditional inputs by changing the production environment (Zhang *et al.* 2012). Following Li *et al.* (2002) and Zhang *et al.* (2012), this study considers R&D as an important environmental variable that may not be capable of producing output directly but is likely to affect the ability of the farm to transform other inputs into outputs more effectively. Although conventional parametric models consider the effects of R&D as a neutral shift variable, the shift of the production function is more likely to be non-neutral. There are some previous studies, for example Swamy (1970) and Kalirajan and Obwona (1995), that apply the varying coefficients regression model to capture the non-neutrality in terms of the observation- and input-specific response coefficients. However, they need restrictive assumptions in estimating their parametric model (Li and Racine 2007).

This paper uses the semi-parametric smooth coefficient model proposed by Hastie and Tibshirani (1993) and Li *et al*. (2002) to investigate the impact of R&D on the productivity of Australia's broadacre farming in a flexible manner. This novel approach accommodates non-neutrality in the effect of R&D on productivity, allowing for varying effects on input elasticities. At the same time, it allows heterogeneities across observations and provides estimates of the marginal effects of

R&D on factor inputs and the output of each firm. Moreover, it estimates both the direct impact of a change in R&D on output and the indirect impact through changes in the efficiency of use of factor inputs in the production process.

The remainder of the paper is organized as follows. Section 2 outlines the econometric methodology, beginning with parametric and Robinson's semi-parametric specifications, followed by the semi-parametric smooth coefficient model. Section 3 describes the data. Section 4 analyses the empirical results. Finally, Section 5 concludes.

2. Methodology: A Semiparametric Smooth Coefficient Model

In the standard literature, firm performance is modelled as a linear function of inputs and other firm level attributes. In practice, the Cobb-Douglas production function, *Model 1*, is perhaps the most widely used parametric regression model in applied research. With all variables measured in logarithms, the production relation being estimated to measure firm performance is:

$$y_i = \alpha_0 + x_i'\beta + z_i\varphi + \epsilon_i \qquad (1)$$

where y is output, x is a vector of firms inputs, z = R&D is the firm's research and development expenditure, β is a vector of unknown parameters and ϵ_i is the identically and independently distributed error term. The ordinary least squares method can then be used to estimate the unknown parameters in Equation 1.

There are, however, two major shortcomings with the standard Cobb-Douglas production function. First, it is necessary to specify the exact parametric form prior to estimation. Hence, it is likely that the presumed model may not be consistent and the inference may not be valid when the model is not correctly specified. In practice, the true parametric form is hardly ever known. Second, in the model, the z variable affects the productivity of all firms in an identical way and constrains the estimation to give constant marginal effects. It does not capture the effects of R&D on individual firms, even though effects may differ across firms and be variable for each firm.

This study considers nonparametric regression methods to address the concern about incorrect parametric specification in the case of modelling inputs and outputs. The Cobb-Douglas functional form is unable to capture the effects of firm characteristics on TFP through the efficiency with which factor inputs are transformed into output. A natural extension of this model that allows the firm characteristics to have a firm-specific effect on TFP is a semi-parametric model. Recently, semi-parametric estimation techniques have drawn much attention among econometricians in the study of firm productivity and efficiency.

This study uses Robinson's (1988) semi-parametric partial linear model, denoted as *Model 2*, to extend the conventional production function with outputs and inputs measured in logarithms as follows:

$$y_i = \alpha(z_i) + x_i'\beta + \epsilon_i \qquad (2)$$

where x_i is a vector of inputs, β is a vector of unknown parameters, and z_i is a vector of environmental variables that enter the model nonlinearly. The functional form of $\alpha(\cdot)$ is not specified and constitutes the nonparametric part of the model. Semi-parametric models are a compromise between fully nonparametric and fully parametric specifications and, thus, are formed by combining parametric and nonparametric models. This specification is in line with the TFP model used in Griffith *et al*. (2004) where $\alpha(z_i)$ is regarded as TFP. The environmental variable, R&D, allows TFP growth to be affected in a flexible way without assuming any particular functional form of z_i variables.

To estimate coefficients in the Robinson model, the basic idea is to first eliminate the unknown function $\alpha(\cdot)$. Taking expectations conditional on z_i for both sides of (2),

$$E(y_i|z_i) = \alpha(z_i) + E(x_i|z_i)'\beta + E(\epsilon_i|z_i)$$

Subtracting this expression from (2) and assuming $E(\epsilon_i|z_i) = 0$ yields;

$$y_i - E(y_i|z_i) = (x_i - E(x_i|z_i))'\beta + \epsilon_i$$

In shorthand notation,

$$\tilde{y}_i = \tilde{x}_i{}'\beta + \epsilon_i$$

Now, β can be estimated by applying the method of least squares:

$$\hat{\beta} = \left[\sum_{i=1}^{n} \tilde{x}_i \tilde{x}_i{}'\right]^{-1} \sum_{1}^{n} \tilde{x}_i \, \tilde{y}_i$$

where $\hat{\beta}$ depends on unknown moments $E(y_i|z_i)$ and $E(x_i|z_i)$ which can be estimated using a nonparametric regression method. Then, replacing them in the above equation yields consistent estimates of $\hat{\beta}$ without modelling $\alpha(z_i)$ explicitly. Finally, $\alpha(z_i)$ can be estimated nonparametrically by regressing $(y_i - x_i'\hat{\beta})$ on z_i. Although the Robinson model is widely used in applied settings and tends to be simpler to interpret than fully nonparametric models, it partially relies on parametric assumptions, and thus, the concerns regarding misspecification and inconsistency are as pertinent for this model as they are for parametric models.

Robinson's (1988) semi-parametric partial linear model introduces the z_i vector into the regression analysis in a fully flexible manner to explain TFP growth. However, this model only allows the R&D variable to have a neutral effect on the production function, that is, it only shifts the level of the production frontier and does not affect the marginal productivity of inputs. In other words, this semi-parametric model does not consider indirect effects of the R&D variable through factor productivity (independent of X variables). Moreover, because it partly depends on parametric assumptions, the issue of misspecification and inconsistency are still relevant.

This study also considers a more general semi-parametric regression model, namely, the semi-parametric smooth coefficient model proposed by Hastie and Tibshirani (1993) and Li *et al.* (2002). Some studies, such as those of Ahmad *et al.* (2005) and Zhang *et al.* (2012), have applied a similar methodology in their productivity analysis in industrial sectors. The semi-parametric smooth coefficient model, *Model 3*, is given by

$$y_i = \alpha(z_i) + x_i'\beta(z_i) + \epsilon_i \qquad (3)$$

where both $\alpha(z_i)$ and $\beta(z_i)$ denote vectors of unspecified smooth functions of z_i. This is one of the most flexible models, and it nests a linear model and a partially linear model (Robinson's semi-parametric model) as special cases. When $\beta(z) = \beta$, this model collapses to the semi-parametric partially linear model, and for a given level of an R&D variable (i.e., when $\beta(z) = \beta$ and $\alpha(z) = \alpha_0$), the semi-parametric smooth coefficient model reduces to constant coefficient parametric Cobb-Douglas functional form (Hartarska *et al*. 2011; Li and Racine 2007).

Specifying input coefficients as unknown smooth functions of z_i, this semi-parametric smooth coefficient model allows indirect effects of the z variable via the input elasticities. For example, if labour and capital are conventional inputs and z_i (R&D expenditures) is an environmental variable, then *Model 3* suggests that the input coefficients of labour and capital may directly vary with firm's R&D. Thus, this model proposes that the marginal productivity of each input, say labour and capital, depends on the firm's z_i variables, such as R&D.

In addition, this generalized model considers the non-neutral impact of R&D on output, capturing the direct effect of z_i variables on TFP and the indirect effects through the efficiency with which factor inputs are used. Furthermore, it provides greater flexibility in the functional form than a parametric linear model or a semi-parametric partially linear model. This functional flexibility allows the model to address the non-neutrality in the production function, which has plagued many applied studies in the past (Li and Racine 2007, 2010). Furthermore, it does not require a sample size as large as that for a nonparametric model. *Model 3* can be expressed more compactly as

$$y_i = \alpha(z_i) + x_i'\beta(z_i) + \epsilon_i = (1, x_i')\begin{pmatrix}\alpha(z_i)\\ \beta(z_i)\end{pmatrix} + \epsilon_i \equiv X_i'\delta(z_i) + \epsilon_i \quad (4)$$

Pre-multiplying (4) by X_i and taking expectations conditional on z_i yields

$$E(X_i y_i | z_i) = E(X_i X_i' | z_i)\delta(z_i)$$

Assuming $E(X_i\epsilon_i|z_i) = 0$ and following Li *et al*. (2002) and Li and Racine (2010), the kernel method can be employed to estimate the following locally constant least squares estimator for $\delta(z)$ as

$$\hat{\delta}(z) = \left(\sum_{j=1}^{n} X_j X_j' K\left(\frac{z_j - z}{h}\right)\right)^{-1} \times \left\{\sum_{j=1}^{n} X_j y_j K\left(\frac{z_i - z}{h}\right)\right\} \quad (5)$$

where $K(\cdot)$ is a kernel function; h is a smoothing parameter or bandwidth, which can be selected via the least squares cross validation method (Li and Racine, 2007); and z_i is the datum at which the kernel function is evaluated. The semi-parametric varying coefficient model has the advantage that it allows greater flexibility in functional forms than a parametric linear model or a semi-parametric partially linear model. At the same time, it avoids much of the "curse of dimensionality" problem (Ahmad *et al*., 2005).

3. Data

This study uses state-level agricultural input and output data collected from annual farm surveys provided by ABARES (Australian Bureau of Agricultural and Resource Economics and Sciences) for the period 1995-2007. The dataset consists of observations on quantities of agricultural inputs, outputs and values of each state for every year during the period. Four major inputs are used: land, labour, capital, and materials. The aggregate value of agricultural production of broadacre agriculture is the measure of output. Data on public investment in agricultural R&D is obtained from Professor John Mullen, who derives the data from the Australian Bureau of Statistics' (ABS) biannual Australian Research and Innovation surveys.[1] The R&D expenditure in broadacre agriculture alone is calculated based on broadacre agriculture's share in the total value of agricultural production.

All estimates except R&D are state-level per farm averages, and all financial estimates are expressed in 2011–2012 Australian dollars as per data sources from AgSurf.[2] In the dataset, *Land* includes all land areas in hectares operated on 30 June by the farm. *Labour* represents the total number of weeks worked by all farm workers, including hired labour. *Capital* includes the value of all assets used on the

[1] Public agricultural R&D includes expenditure by Australian, state and territory governments as well as research institutions and universities. Funds from research and development corporations (excluding grower levies) and other external funders for agriculture (excluding research in fisheries and forestry) are also included.

[2] AgSurf reports state-level per farm average data from the Australian agricultural and grazing industries survey (AAGIS) and Australian dairy industry survey (ADIS) conducted by ABARES

farm, including leased equipment but excluding machinery and equipment either hired or used by contractors. ABARES uses the market value of livestock/crop inventories and replacement value less depreciation for plants and machinery in calculating the value of capital. *Materials* includes farm expenditures on seeds, crop and pasture chemicals, fuel oil and grease, livestock materials, contracts (cropping and livestock), fertilizer, shearing crutching and other materials and services. The final sample includes 65 observations (5 states over 13 years) with complete records for the variables mentioned above.

Studies suggest that there is a lag relationship between R&D and productivity growth, and a credible estimate of the effects of R&D on subsequent productivity relies on specifying the lag structure (Griliches 1998). There are various lag structures used in studies in estimating the impacts of R&D expenditure on productivity, which may vary between 10 to 30 years to approximate the right lag structure. However, the short data series restricts us from directly modelling the length and shape of the R&D lag in this study. As one of the simplest ways of accommodating the lag structure in empirical studies, TFP is specified as a function of knowledge stocks, which are determined by current and past R&D expenditures (Griliches 1979; Thirtle and Schimmelpfennig 2008). This thesis constructs a simple R&D knowledge stock variable using a perpetual inventory model (PIM). In this method, R&D stocks are calculated from flow of R&D expenditures based on the following equation:

$$K_t = R\&D_t + (1 - \delta)K_{t-1} \qquad (6)$$

where K_t is the R&D knowledge stock at time t, $R\&D_t$ is the agricultural R&D expenditure at the time t and δ is the depreciation rate for R&D knowledge stock.

The initial stock is calculated as:

$$K_0 = \frac{R\&D_0}{g + \delta}$$

where $R\&D_0$ is the R&D expenditure in the first year available, and g is the average annual logarithmic growth of R&D expenditure for every state over the period of analysis. This PIM method is used as a simple alternative to a complex time-lag structure between current productivity and the flow of past R&D investments.

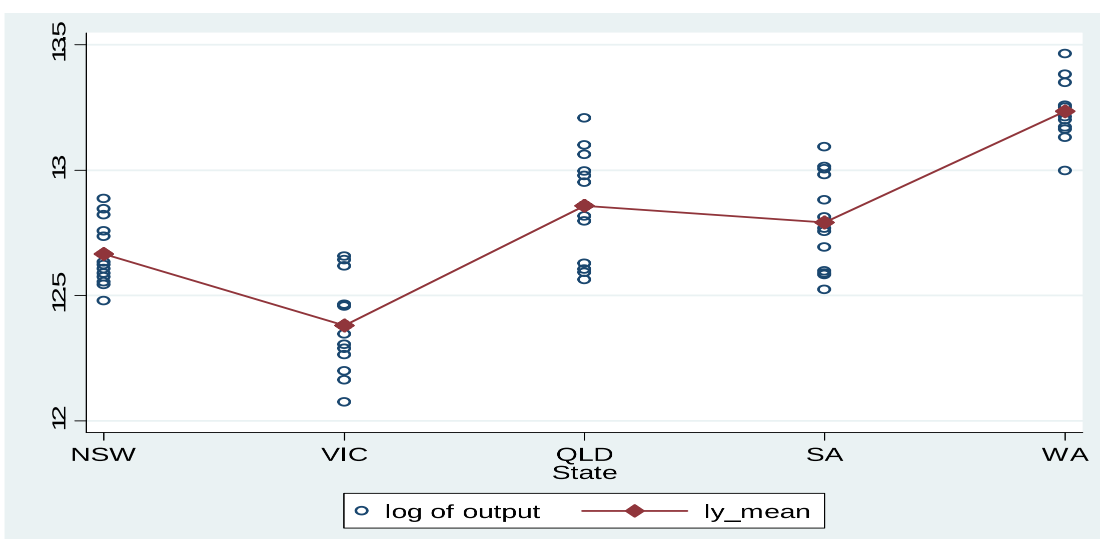
However, a limitation of the PIM method is the need to choose a depreciation rate, which varies within the range 0.05 to 0.10 across econometric studies in agriculture (Thirtle *et al*. 2008). This research sets a depreciation rate of R&D fixed at 8 per cent. Table 1 reports the summary statistics for the natural logarithms of the variables. Figure 1 shows state-level heterogeneity in output means. As can be seen from the figure, there are large variations in terms of output across states.

Table 1: Summary statistics

Variable	Obs	Mean	Standard deviation	Minimum	Maximum
ln Output	65	12.7858	0.32580	12.07448	13.46515
ln Capital	65	14.6511	0.37637	14.05605	15.52822
ln Labour	65	4.62423	0.13361	4.35671	4.89035
ln Land	65	8.34441	1.12009	6.40853	9.60407
ln Materials	65	11.0072	0.37565	10.30189	12.08648
ln R&D	65	14.4151	0.87196	12.98421	15.68413

Source: Authors' own calculations

Figure 1: State-level output variations



4. Empirical Results

In this section, results are presented from the different production function specifications mentioned in the methodology section, starting with a simple Cobb-Douglas model and generalizing it stepwise through a semi-parametric partial linear model and a semi-parametric smooth coefficient model. These models are nested, which means that the semi-parametric smooth coefficient model can reduce with appropriate restrictions to the traditional Cobb-Douglas production model with constant elasticities. Hence, the specifications can be tested against each other.

Table 2 shows the results from Models 1, 2 and 3. Model 1 is a simple Cobb-Douglas production function where (log) output is modelled as a linear function of (log) factor inputs and is extended to include an environmental variable, (log) *R&D* investment. The *R&D* variable is introduced additively and parametrically and the model is estimated using OLS. The estimates of the conventional Cobb-Douglas production specifications are reported in column 2 under the heading *Model 1*. The results show that the estimated coefficients of two major inputs, *capital* and *labour,* are both positive and significant. The estimated coefficient of R&D captures the marginal effect of *R&D* on productivity, which is constrained to be the same across the states. The results do not suggest R&D has a significant influence on productivity growth.

Model 2, which estimates Robinson's semi-parametric partial linear model is used to bring flexibility into the specification. It allows the effects of R&D in a flexible manner and captures the state-specific impact of the R&D variable on productivity through TFP.[3] In this model, (log) output is modelled as a linear function of (log) factor inputs as in Model 1, and the R&D variable enters the model nonparametrically by introducing the intercept term as an unknown (flexible) function of the R&D variable. Model 2 captures the non-linearity in the relation between the output and R&D. The estimates of the semi-parametric partially linear model are presented in column 3 in Table 2, which shows that the coefficients of the

[3] Model 2 is estimated with *semipar* package of STATA software where the variable R&D enters the model nonlinearly. The Gaussian kernel function is used to estimate the regressions nonparametrically in a local weighted polynomial fit. In addition, in Model 2 intercept term could not be identified separately from the unknown function $\alpha(\cdot)$.

capital and labour inputs are positive and significant, as in Model 1. In addition, it shows a negative but insignificant partial effect of the environmental variable R&D.

Table 2: Parametric and semi-parametric regression coefficients: pooled data

Variables	Model 1	Model 2	Model 3
	OLS	Robinson's Semi-parametric	Semi-parametric smooth coefficients
Capital	0.251**	0.290***	0.3136***
	(0.113)	(0.0790)	(0.1144)
Labor	1.254***	0.664**	0.8298**
	(0.315)	(0.303)	(0.1477)
Land	0.0152	-0.0849**	0.0988
	(0.0361)	(0.0368)	(0.0884)
Materials	0.119	-0.0386	0.00832
	(0.130)	(0.108)	(0.1267)
R&D	-0.0696	-0.1153	0.0653*
	(0.0443)	(.0717)	(0.0386)
Constant	2.879**		3.406**
	(1.190)		(0.5530)
Observations	65	65	65
R-squared	0.801	0.416	0.9303

Robust standard errors in parentheses in Model 1 & 2. Model 3 reports bootstrapped standard errors. *** p<0.01, ** p<0.05, * p<0.1

Finally, *Model 3*, which is termed the semi-parametric smooth coefficient model, brings more flexibility in the specifications where both intercept and input coefficients are unknown, and it provides a smooth function of the environmental variable R&D. In both *Models 1 and 2*, R&D shifts the production frontier neutrally, i.e., the input elasticities are invariant with respect to R&D, although in *Model 2*, R&D allows TFP growth to be affected in a flexible manner.

In *Model 3*, R&D is allowed to non-neutrally affect the production function, where both the intercept and slope coefficients are modelled as an unknown smooth function of the R&D variable.[4] Thus, in this model, the input coefficients, i.e., the

[4] Model 3 is computed using the *np* package of the R software (*version 3.1.0 "Spring Dance"*). The smooth coefficient Kernel Regression *npscoef* functions is used with the bandwidth selection

elasticity of output with respect to capital, labour, land, and materials are allowed to vary with respect to R&D. This model is estimated nonparametrically using the semi-parametric smooth coefficient model proposed by Li *et al*. (2002) and Li and Racine (2010), where the local constant least squares procedure is applied to estimate these functional coefficients. Table 2 reports the mean values for *Model 3*, as it gives rise to observation-specific estimates (detailed results of *Model 3* are reported in Table 3).

There are some variations in terms of magnitude, sign and significance across the three different models presented in Table 2. The elasticities of output with respect to the capital ($\hat{\beta}_1$) and labour ($\hat{\beta}_2$) inputs are positive and significant across each of the three specifications. The marginal effect of R&D on output is positive and significant only in *Model 3*. The negative effects of R&D in both the Cobb-Douglas parametric model (*Model 1*) and Robinson's semi-parametric model (*Model 2*), though insignificant, are inconsistent with conventional expectations.

Model 3 as a local-linear regression follows the rule-of-thumb that the bandwidth needs to be less than twice the standard deviation (σ_z) of the continuous variable to enter the model non-linearly.[5] This implies that the R&D variable does not enter the model in the linearly and additively separate fashion assumed in the conventional parametric specification – *Model 1*. These statistical results are economically meaningful and make the semi-parametric smooth coefficient model (*Model 3*) more appealing than the corresponding parametric model or Robinson's semi-parametric model.

Table 3 summarizes the detailed results from the semi-parametric smooth coefficient production specification - *Model 3*. Because *Model 3* gives observation-specific estimates, the summary results are reported at the mean, 1st Quartile (25th percentile), Median (2nd Quartile), and 3rd Quartile (75th percentile), along with minimum and maximum values. The results show a large variation in the marginal impacts of environmental variable R&D on farm performance in the semi-parametric smooth coefficient model. This heterogeneity in impact suggests that the traditional Cobb-

npscoefbw function, *bwmethod* = "*cv.ls*" (least squares cross validation), *ckertype* (continuous kernel type) = "gaussian". Semi-parametric fits of the estimates are obtained using the bootstrapped standard error.

[5] $2 \times \sigma_z = 1.7438\ and\ Bandwidth\ for\ Z\ variable = 0.2057$.

Douglas production model capturing the average (or mean) impact of the R&D variable is not appropriate. The marginal effects of R&D on the elasticities of the factor inputs at the mean and at each of the three quartile values suggest that impact of R&D on production technology is not input neutral.

The environmental variable, R&D, affects the marginal productivity of inputs in a non-neutral manner, as indicated in Table 3. It has both a direct effect through TFP ($\partial\hat{\beta}_0/\partial lnZ$) and an indirect effect via the productivity ($\partial\hat{\beta}_i/\partial lnZ$) with which the inputs are used in the production process. The marginal effect of the environmental variable on overall productivity, $\partial lnY/\partial lnZ$, (here Z is R&D) is given by

$$\frac{\partial lnY}{\partial lnZ} = \overbrace{\frac{\partial\hat{\beta}_0}{\partial lnZ}}^{\text{Direct Effect}} + \overbrace{\frac{\partial\hat{\beta}_1}{\partial lnZ}k + \frac{\partial\hat{\beta}_2}{\partial lnZ}l + \frac{\partial\hat{\beta}_3}{\partial lnZ}a + \frac{\partial\hat{\beta}_4}{\partial lnZ}m}^{\text{Indirect Effects}} \qquad (7)$$

where k is (log) capital, l is (log) labour, a is (log) land and m is (log) materials.

The seventh column of Table 3 reports the marginal productivity of R&D (i.e., the elasticity, $\partial lnY/\partial lnZ$). R&D has a positive and statistically significant effect on output with a mean value of 0.0653, which means that for a 1 per cent increase in R&D investment, the output responds positively by 0.0653 per cent, on average.

Table 3: Summary of the results for semi-parametric smooth coefficients

1	2	3	4	5	6	7	8	9	10	11	12
Variable	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\partial lnY/\partial lnZ$	$\partial\hat{\beta}_0/\partial lnZ$	$\partial\hat{\beta}_1/\partial lnZ$	$\partial\hat{\beta}_2/\partial lnZ$	$\partial\hat{\beta}_3/\partial lnZ$	$\partial\hat{\beta}_4/\partial lnZ$
Mean	3.406	0.3136	0.8299	0.0988	0.0083	0.0653	5.264	0.0136	0.0807	-0.1995	0.1555
	(0.5530)	(0.1144)	(0.1477)	(0.0884)	(0.1267)	(0.0386)	(1.2619)	(0.0796)	(0.2752)	(0.1123)	(0.0867)
1st Qu.	2.088	0.0357	0.4866	-0.052	-0.262	0.0435	-1.281	-0.4573	-0.7503	-1.1110	-0.4238
	(0.0528)	(0.0291)	(0.0921)	(0.0382)	(0.0436)	(0.0135)	(0.9957)	(0.0305)	(0.1395)	(0.1030)	(0.0721)
Median	3.351	0.3566	1.1284	0.0802	-0.110	0.0521	4.543	0.1913	0.3342	-0.2201	-0.1123
	(0.2654)	(0.0613)	(0.1806)	(0.0277)	(0.0621)	(0.0081)	(0.9014)	(0.1330)	(0.2270)	(0.0306)	(0.2020)
3rd Qu.	3.809	0.5925	1.5343	0.2994	0.2822	0.0728	11.440	0.4328	1.1600	-0.0664	0.6140
	(1.2196)	(0.0819)	(0.2573)	(0.0177)	(0.0432)	(0.0115)	(1.1288)	(0.0829)	(0.3990)	(0.1965)	(0.0386)
Min	-3.077	-0.188	-0.891	-0.169	-0.406	-0.1084	-10.490	-1.4270	-6.6310	-1.4330	-1.2170
	(0.0537)	(0.0004)	(0.0890)	(0.0169)	(0.0261)	(0.0334)	(1.3571)	(0.2192)	(0.1580)	(0.1375)	(0.4321)
Max.	9.738	0.7167	1.9141	0.3811	0.6059	0.5199	27.510	0.8374	3.4310	0.3567	1.4320
	(1.0442)	(0.0234)	(0.0992)	(0.0206)	(0.0814)	(0.2429)	(7.870)	(0.1580)	(0.6602)	(0.0066)	(0.2425)

Bootstrapped standard errors in parentheses

The results also show that there is some variation in the marginal effects of R&D on overall productivity, with a range of effects from -0.11 per cent to 0.52 per cent. These marginal effects are the combined effect of both direct and indirect effects of R&D on productivity. The results reported in column 8 show substantial heterogeneity in the direct effects of R&D on TFP ($\partial\hat{\beta}_0/\partial lnZ$). This article follows the residual based wild bootstrap method to estimate standard errors in the semi-parametric smooth coefficient model.[6]

The marginal effects of *R&D* on the factor productivity of inputs vary across the inputs as well as over the observations in the sample. On average, the effects of *R&D* on input productivity are 0.0136 per cent, 0.0807 per cent, -0.1995 per cent and 0.1555 per cent for *capital*, *labour*, *land* and *materials*, respectively. These results indicate that all inputs except *land* have positive contributions of R&D to productivity, and the effect is biased towards the increased productivity of *materials*. The greatest variation is found in the marginal effect of *R&D* on the contribution of *labour* to output ($\partial\hat{\beta}_2/\partial lnZ$), with minimum and maximum values of -6.63 per cent and 3.43 per cent, respectively.

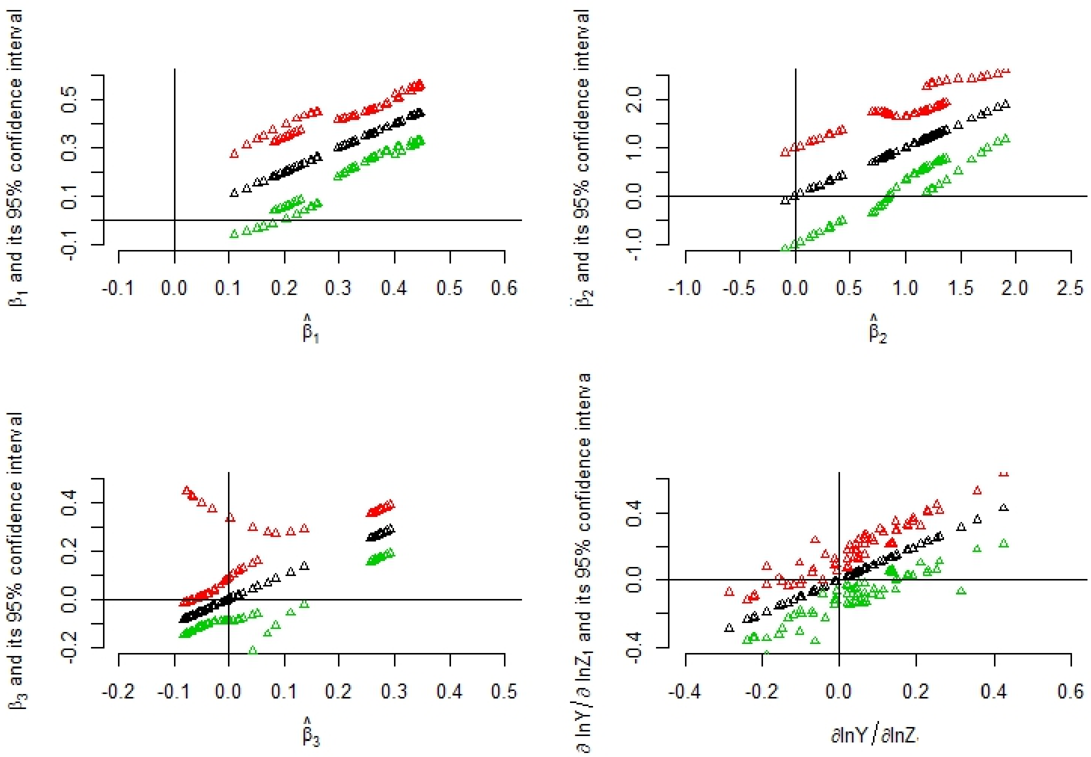
Figure 2 plots the partial effects for each observation in the sample ordered by the value of the estimated coefficient, along with bootstrapped confidence bounds for each of the partial effects. The advantage of this type of plot is that it shows statistical significance for the partial effect of each observation.[7] Here the plot shows substantial heterogeneity in the coefficients of the observation-specific partial effects of capital ($\hat{\beta}_1$), labour ($\hat{\beta}_2$), land ($\hat{\beta}_3$) and R&D ($\partial \ln Y/\partial \ln Z$). For most of the observations the lower bounds of the input coefficients, $\hat{\beta}_1$, $\hat{\beta}_2$, and $\hat{\beta}_3$, are greater than zero, indicating positive and statistically significant estimates of output elasticities with respect to capital, labour and land. Turning to the marginal effects of R&D, $\partial \ln Y/\partial \ln Z$ (where Z is R&D), Figure 2 also shows a plot of the marginal effects of the R&D. It is found that although R&D has both positive and negative effects on output, the effect at the mean is positive and statistically significant. Therefore,

[6] Following steps are followed: (i) Obtain fitted residuals, $\hat{\varepsilon}_i$, from the sample; (ii) Generate wild bootstrap disturbance, ε_i^*, such that the distribution of two points is as follows: $\varepsilon_i^* = a\hat{\varepsilon}_i$ with probability $r = (\sqrt{5}+1)/(2\sqrt{5})$ and $\varepsilon_i^* = b\hat{\varepsilon}_i$ with probability $1-r$, where $a = -(\sqrt{5}-1)/2$ and $b = (\sqrt{5}+1)/2$, as suggested by Mammen (1993); (iii) Resample the response variable y_i^* based on the bootstrapped disturbance, ε_i^*; (iv) Refit the model using the fictitious response variables; and (v) Repeat steps 2 and 4 a statistically significant number of times, say, B=99.

[7] The following procedure is followed to construct these plots. For any given estimate, say, $\hat{\beta}_1$, $\hat{\beta}_1$ is plotted against $\hat{\beta}_1$, which plots $\hat{\beta}_1$ along the 45 degree line. Then, to obtain the confidence bounds the standard error is added (subtracted) twice from $\hat{\beta}_1$, which gives the upper (lower) confidence bounds. The upper and lower confidence bounds are plotted against $\hat{\beta}_1$.

only considering the impact of R&D on the average can be misleading when there is non-neutrality in the effects of R&D investment.

Figure 2: Semi-parametric fits: estimates with confidence intervals



To check the robustness of the estimates, we treat the dataset as a panel (repeated cross section over periods) and use both the fixed effects and the semi-parametric smooth coefficient model. These panel specifications control state-level unobserved fixed effects in analysing the data. Table 4 presents estimates of the parametric (fixed effects) and semi-parametric smooth coefficient models (plot of the marginal effects of R&D based on panel data is reported in appendix Figure A.1). Like the OLS model, the fixed effects model shows that the input coefficients for both capital and labour are positive and significant and that the R&D coefficient is negative but insignificant. In turn, the partial effect of R&D is positive and significant for the semi-parametric smooth coefficient model with panel data. These results suggest the estimates are robust for panel data as well.

Table 4: Fixed effects and semi-parametric smooth coefficients: panel data

Variables	Fixed Effects Model	Semi-parametric smooth coefficient model
Capital	0.346*	0.2889**
	(0.162)	(0.0886)
Labor	0.856*	0.8525**
	(0.323)	(0.284)
Land	0.0457	0.1433
	(0.214)	(0.0834)
Materials	-0.0590	0.1154
	(0.176)	(0.0844)
R&D	-0.140	0.1024*
	(0.182)	(0.0507)
Constant	7.477**	1.0497
	(2.230)	(1.5153)
Observations	65	65
R-squared	0.590	0.9048

Robust standard errors in parentheses

**** p<0.01, ** p<0.05, * p<0.1*

Several model specification tests are applied to formally test for the correct specification. Firstly, to test the parametric specification of *Model 1*, the Ramsey RESET model specification test is used. The test using powers of the independent variables produces a significant test statistic $F\ (12, 48) = 4.00$ with $Prob > F = 0.0003$ for specification error. This test suggests rejection of the null hypothesis that the model has no omitted variables, and it indicates that the parametric specification is not a correct specification. Secondly, Hardle and Mammen (1993) specification test is implemented to check whether the nonparametric fit can be approximated by a polynomial fit in any order. Absence of rejection of the null (i.e., "accepting" the parametric model) means that the polynomial adjustment is at least of the degree that has been tested. The test statistics reported in Table 5 show that the parametric model could be approximated with a polynomial fit of degree 3 of R&D.

[8] The pseudo R-squared is derived as the square of the Pearson product moment correlation coefficient, r. This correlation coefficient is based on the correlation between the predicted values and the actual values in the model, which can range from -1 to 1, and so the square of the correlation then ranges from 0 to 1.

Table 5: Hardle and Mammen's (1993) specification test

H0: Parametric and nonparametric fits are not different		
Polynomial Degree	Approximate P-value	Decision
1	0.0	Linear approximation rejected
2	0.07	Quadratic approximation rejected
3	0.21	Cubic approximation cannot be rejected

Thirdly, to choose the preferred model, we also use the Cai, Fan and Yao model specification test proposed by Cai *et al.* (2000). This test is used to determine which model best fits the data between the parametric and smooth semi-parametric models. This test is based on a comparison of the residual sum of squares (RSS) from both parametric and semi-parametric fittings. The test statistic is defined as

$$T_n = \frac{\left(RSS_{para} - RSS_{semipara}\right)}{RSS_{semipara}} = \frac{RSS_{para}}{RSS_{semipara}} - 1$$

where a large value of T_n suggests rejection of the null hypothesis. A nonparametric bootstrap approach is used to evaluate the p-value of the test. The bootstrapped test statistic T_n^* is calculated from the generated bootstrap residuals from the semi-parametric fit. The p-value of the test is simply the relative frequency of the event $T_n^* \geq T_n$ in the bootstrap samples. The goodness-of-fit test statistics suggest rejecting the null hypothesis that both the parametric and nonparametric fittings are the same with a p-value equal to 0.00. Hence, the semi-parametric smooth coefficient model is the preferred specification in this case. This result confirms that the production function is of the variable coefficient type and that the impact of R&D on output is non-neutral and input specific. Therefore, the semi-parametric smooth coefficient model is more appealing because of its ability to capture both direct and indirect effects of the environmental variable, R&D.

Finally, a likelihood ratio test is also performed for adding a time variable to the model. The test gives the likelihood ratio test statistic, a chi-square of 0.53 with one degree of freedom, as well as the associated p-value of 0.4658 (*LR chi2(1 = 0.53; Prob > chi2 = 0.4658*). Thus, according to the data, it cannot be rejected the null hypothesis that the model excludes a time variable. The results show that adding time as a predictor variable does not result in a statistically significant improvement in model fit. Moreover, an F test is performed to see if time-fixed effects are needed when running a fixed effect model. The null is that no time-fixed effects are needed and that all-year dummies are jointly non-significant. In the case of our sample, the test statistic is 1.55 with a p-value 0.3404 ($F\ (4,\ 4) = 1.55;\ Prob > F = 0.3404$). This indicates that the sample data are compatible with the null hypothesis that no time-fixed effects are needed in the model.

These specification tests generally reject the parametric specifications in favour of more flexible counterparts. This result is consistent with studies that apply similar methodologies but perform the tests in the manufacturing sector. For example, Li *et al*. (2002) use the nonparametric kernel method to estimate the semi-parametric varying coefficient model with China's non-metal mineral manufacturing industry data. They find that the semi-parametric varying coefficient model is more appropriate than either a parametric linear model or a semi-parametric partially linear model. Similarly, using a provincial-level dataset Zhang *et al*. (2012) suggest that the semi-parametric model yields outcomes that are more intuitive and have fewer economic violations than the parametric counterpart in China's high technology industry.

5. Conclusion

The conventional econometric approaches ordinarily produce point estimates of the effect of R&D on the productivity of the average unit of analysis assuming implicitly that environmental variables influence productivity neutrally, through the TFP alone, and the differential effect of R&D on factor inputs is not recognized. As a result, the policy implications for R&D investment turn into a one-size-fits-all sort of strategy. Against this backdrop, this article uses a novel econometric methodology, the semi-parametric smooth coefficient model to analyse the effect of R&D on productivity in Australian broadacre farming. This approach gives rise to the observation-specific estimates of input coefficients. Using state-level average farm data, estimates are provided of the state-level effect of R&D on productivity and the marginal productivity with which factor inputs are used in the production process.

The results show that the R&D does not have the same effect on TFP and productivity at the average farm level across states within Australia. By specifying intercept and slope coefficients as a function of the environmental variable, R&D, the model gives rise to significant variation in the state-level effects of R&D. Therefore, this study confirms the non-neutrality in the effects of R&D on productivity. The estimates of the effect of R&D

investments on productivity in broadacre farming are more useful than that of parametric estimates in terms of policy implications. First, the results suggest that Australia may enhance its farming productivity by improving investment in public R&D. Second, the large variations in the state-level average farm effects of R&D on productivity imply that initiation of the same R&D policy in different states can have considerably diverse effects on the productivity of inputs. Furthermore, R&D expenditure is found to have a direct impact on productivity and indirect effects through impacting the marginal productivity of factor inputs such as labour and capital. Importantly, none of these issues come into consideration in the parametric regression specifications of modelling the impact of R&D on productivity. This is the fundamental point of interest of using this novel methodology.

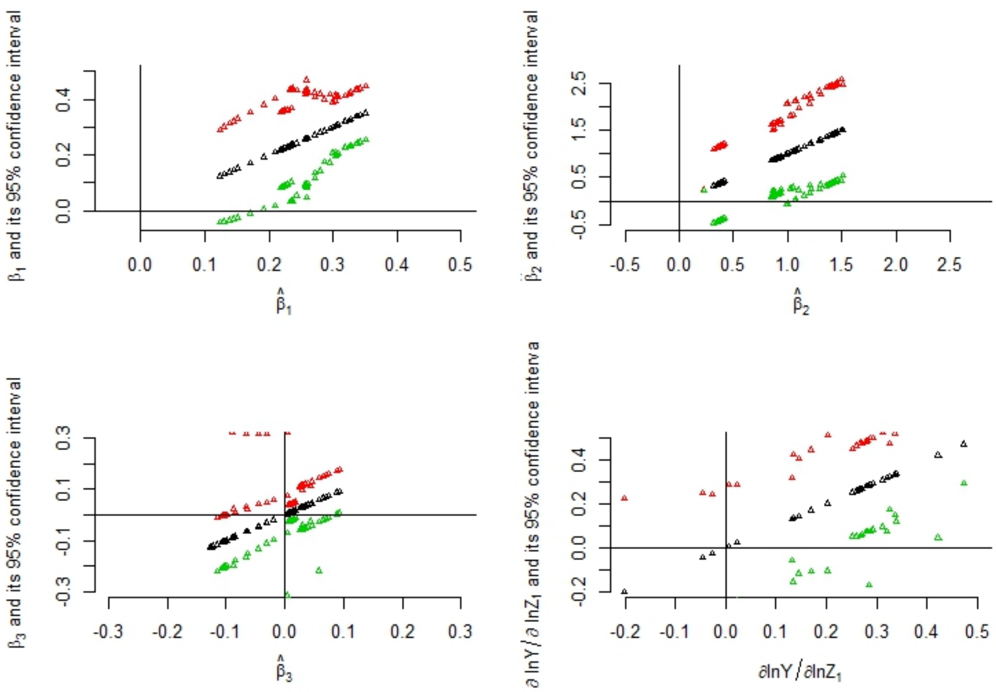
Finally, the results provide evidence that the effect of environmental variables on economic performance needs to be revisited. Specifically, consideration should be given to the variations in the effect of R&D on farms. This study has limitations in that it could not consider the effect of private R&D due to data unavailability. However, other studies show that increased spending on public research appears to supplement private research in agriculture (Wang et al., 2013). Another limitation is that the within-state variations in the effects of R&D are not estimated, as data are available only at the aggregate state level. In addition, the possibility of errors of measurement with the state-level public R&D data cannot be ruled out. Nevertheless, this research explores the relationship from a novel methodological point of view and broadly confirms the results of previous studies regarding the average impact of R&D on productivity, and it provides the additional insight that R&D affects productivity non-neutrally and differentially across farms.

Acknowledgement

This study acknowledges Professor Subal C. Kumbhakar (Binghamton University, USA), Professor Jeffrey Racine (McMaster University, Canada) and Dr Kai Sun (Aston University, UK) for their advice and materials used in estimating the semi-parametric smooth coefficient model. We are thankful to the participants and discussant in the 2014 Asia-Pacific Productivity Conference at the University of Queensland and the 2014 PhD Conference in Economics and Business at Monash University for generous inputs.

Appendix

Figure A.1: Semi parametric fits with panel data



References

Ahmad I, Leelahanon S, Li Q (2005) Efficient Estimation of a Semiparametric Partially Linear Varying Coefficient Model. Annals of Statistics 33(1): 258-283

Alene AD (2010) Productivity Growth and the Effects of R&D in African Agriculture. Agricultural Economics 41(3-4): 223-238

Alston JM, Andersen MA, James JS, Pardey PG (2011) The Economic Returns to US Public Agricultural Research. American Journal of Agricultural Economics 93(5): 1257-1277

Alston JM, Pardey PG (2014) Agriculture in the Global Economy. The Journal of Economic Perspectives 28(1): 121-146

Ball E, Schimmelpfennig D, Wang SL (2013) Is U.S. Agricultural Productivity Growth Slowing? Applied Economic Perspectives and Policy 35(3): 435-450

Bervejillo JE, Alston JM, Tumber KP (2012) The Benefits from Public Agricultural Research in Uruguay. Australian Journal of Agricultural and Resource Economics 56(4): 475-497

Cai Z, Fan J, Yao Q (2000) Functional-Coefficient Regression Models for Nonlinear Time Series. Journal of American Statistical Association 95(451): 941-956

Coe DT, Helpman E (1995) International R&D Spillovers. European Economic Review 39(5): 859-887

Fuglie KO, Toole, AA (2014) The Evolving Institutional Structure of Public and Private Agricultural Research. American Journal of Agricultural Economics. doi: 10.1093/ajae/aat107

Griffith R, Redding S, Van Reenen J (2004) Mapping the Two Faces of R&D: Productivity Growth in a Panel of OECD Industries. Review of Economics and Statistics 86(4): 883-895

Griliches Z (1964) Research Expenditures, Education, and the Aggregate Agricultural Production Function. American Economic Review 54(6): 961-974

---------- (1979) Issues in Assessing the Contribution of Research and Development to Productivity Growth. Bell Journal of Economics 10(1): 92-116

---------- (1998) Introduction to "R&D and Productivity: The Econometric Evidence". In R&D and Productivity: The Econometric Evidence. University of Chicago Press: 1-14

Hardle W, Mammen E (1993) Comparing Nonparametric versus Parametric Regression Fits. Annals of Statistics 21(4): 1926-1947

Hartarska V, Parmeter CF, Nadolnyak D (2011) Economies of Scope of Lending and Mobilizing Deposits in Microfinance Institutions: A Semiparametric Analysis. American Journal of Agricultural Economics 93(2): 389-398

Hastie T, Tibshirani R (1993) Varying-coefficient Models. Journal of the Royal Statistical Society. Series B (Methodological) 55(4): 757-796

Islam N, Xayavong V, Kingwell R (2014) Broadacre Farm Productivity and Profitability in South-Western Australia. Australian Journal of Agricultural and Resource Economics 58(2): 147–170

Kalirajan KP, Obwona M B (1995) Estimating Indivudual Response Coefficients in Varying Coefficients Regression Models. Journal of Applied Statistics 22(4): 477–484

Khan F, Salim R, Bloch H (2014) Nonparametric Estimates of Productivity and Efficiency Change in Australian Broadacre Agriculture. Australian Journal of Agricultural and Resource Economics. doi: 10.1111/1467-8489.12076

Li Q, Huang CJ, Li D, Fu TT (2002) Semiparametric Smooth Coefficient Models. Journal of Business & Economic Statistics 20(3): 412-422

Li Q, Racine JS (2007) Nonparametric Econometrics: Theory and Practice. Princeton University Press: 301-312

---------- (2010) Smooth Varying-Coefficient Estimation and Inference for Qualitative and Quantitative Data. Econometric Theory 26(6): 1607-1637

Mamuneas TP, Savvides A, Stengos T (2006) Economic Development and the Return to Human Capital: A Smooth Coefficient Semiparametric Approach. Journal of Applied Econometrics 21(1): 111-132

Mullen J (2007) Productivity Growth and the Returns from Public Investment in R&D in Australian Broadacre Agriculture. The Australian Journal of Agricultural and Resource Economics 51(4): 359-384

---------- (2010) Trends in Investment in Agricultural R&D in Australia and its Potential Contribution to Productivity. Australasian Agribusiness Review 18(2): 18-29

Mullen JD, Scobie GM, Crean J (2008) Agricultural Research: Implications for Productivity in New Zealand and Australia. New Zealand Economic Papers 42(2): 191-211

Parmeter CF, Sun K, Henderson DJ, Kumbhakar SC (2014) Estimation and inference under economic restrictions. Journal of productivity analysis 41(1): 111-129

Pardey PG, Alston JM, Chan-Kang C (2013) Public Agricultural R&D Over the Past Half Century: An Emerging New World Order. Agricultural Economics 44(S1): 103-113

Piesse J, Thirtle C (2010) Agricultural R&D, Technology and Productivity. Philosophical Transactions of the Royal Society of London. Series B, Biological sciences 365(1554): 3035-3047

Rahman S, Salim R (2013) Six Decades of Total Factor Productivity Change and Sources of Growth in Bangladesh Agriculture (1948–2008). Journal of Agricultural Economics 64(2): 275-294

Robinson PM (1988) Root-N-Consistent Semiparametric Regression. Econometrica 56(4): 931-954

Salim R, Islam N (2010) Exploring the Impact of R&D and Climate Change on Agricultural Productivity Growth: The Case of Western Australia. Australian Journal of Agricultural and Resource Economics 54(4): 561-582

Schultz TW (1953) The Economic Organization of Agriculture. Bombay: Tata McGraw-Hill

Sheng Y, Zhao S, Nossal K, Zhang D (2014) Productivity and Farm Size in Australian Agriculture: Reinvestigating the Returns to Scale. Australian Journal of Agricultural and Resource Economics 59(1): 16-38

Sun K, Kumbhakar SC (2013) Semiparametric Smooth-coefficient Stochastic Frontier Model. Economics Letters 120(2): 305-309

Suphannachart W, Warr P (2011) Research and Productivity in Thai Agriculture. Australian Journal of Agricultural and Resource Economics 55(1): 35-52

Swamy PAVB (1970) Efficient Inference in a Random Coefficient Regression Model. Econometrica 38(2): 311-323

Thirtle C, Piesse J, Schimmelpfennig D (2008) Modeling the Length and Shape of the R&D Lag: An Application to UK Agricultural Productivity. Agricultural Economics 39: 73–85

Voutsinas I, Tsamadias C (2014) Does Research and Development Capital Affect Total Factor Productivity? Evidence from Greece. Economics of Innovation and New Technology 23(7): 631-651

Wang SL, Heisey PW, Huffman WE, Fuglie KO (2013) Public R&D, Private R&D, and U.S. Agricultural Productivity Growth: Dynamic and Long-Run Relationships. American Journal of Agricultural Economics 95(5): 1287-1293

Zhang R, Sun K, Delgado MS, Kumbhakar SC (2012) Productivity in China's High Technology Industry: Regional Heterogeneity and R&D. Technological Forecasting & Social Change 79(1): 127-141

Zhao W, Zhang R, Liu J, Lv Y (2014) Robust and Efficient Variable Selection for Semiparametric Partially Linear Varying Coefficient Model based on Modal Regression. Annals of the Institute of Statistical Mathematics 66(1): 165-191