

Test of Convergence in Agricultural Factor Productivity: A Semiparametric Approach

Krishna P. Paudel, Louisiana State University and LSU Agricultural Center

Mahesh Pandit, Louisiana State University and LSU Agricultural Center

Biswo N Poudel, University of California Berkeley

Contact Information

Krishna P. Paudel

Associate Professor

Department of Agricultural Economics and Agribusiness

225 Martin D. Woodin Hall

Louisiana State University and LSU Agricultural Center

Baton Rouge, LA 70803:

Phone: (225) 578-7363

Fax: (225) 578 -2716

Email: kpaudel@agcenter.lsu.edu

Selected Paper prepared for presentation at the Agricultural & Applied Economics Association's 2011 AAEA & NAREA Joint Annual Meeting, Pittsburgh, Pennsylvania, July 24-26, 2011

Copyright 2011 by [Krishna Paudel]. All rights reserved. Readers may make verbatim copies of this document for non-commercial purposes by any means, provided that this copyright notice appears on all such copies.

Test of Convergence in Agricultural Factor Productivity: A Semiparametric Approach

Abstract

We tested for club convergence in U.S. agricultural total factor productivity using a sigma convergence test. We used the same club of states as used by McCunn and Huffman as well as different states within 10 clubs identified by the cluster analysis. Results showed convergence was evident only in a few club groups. Clusters group identified using a statistical method identified only converging clubs. Variables affecting total factor productivity among states were identified using parametric, semiparametric and nonparametric methods. Semiparametric and nonparametric methods gave a better fit than a parametric method as indicated by the specification test. Our results indicated that health care expenditure, public research and extension investment, and private expenditure are important variables impacting total factor productivity differences across states.

Keywords: clubs, sigma convergence, cluster analysis, semiparametric and nonparametric methods

Test of Convergence in Agricultural Factor Productivity: A Semiparametric Approach

Exploration on the idea of convergence in agricultural productivity has been an active topic of research (Ball, Hallahan, and Nehring 2004; McCunn and Huffman 2000; Mukherjee and Kuroda 2003; Liu et al. 2010; Poudel et al. 2011; Rezitis 2005, 2010; Thirtle et al. 2003). We propose to test the presence of convergence in U.S. agricultural total factor productivity using recently available data from the ERS/USDA. We also develop a semiparametric model to explain the role of different variables on productivity differences across states using panel parametric and semiparametric methods. Agricultural productivity has been increasing in the U.S. states for several decades although this gain is not the same across states. If productivity is different across states, for policy reasons, it is important to identify factors contributing to this difference in total agriculture factor productivity.

There are two types of convergence: β -convergence and σ -convergence. β -convergence refers to situation where a state with lower total factor productivity at the beginning grows faster than the other states. σ -convergence refers to situation where dispersion across a group of countries decreases over time. Convergence can be tested using many types of tests (time series test such as using unit root, panel data based test). When testing for a convergence the general tendency is to test for absolute or regional convergence. Previous productivity convergence test in the U.S. has adopted grouping based on weather (McCunn and Huffman, 2000), and identification of grouping of states using a cluster based approach (Poudel et al. 2011). We also use a cluster based approach which is in a way similar to Philips and Sul's (2007) convergence test.

Agricultural total factor productivity in the U.S. does not show an absolute convergence as found by the previous authors (McCunn and Huffman 2000; Poudel et al. 2011). These authors have found the presence of club level convergence. McCunn and Huffman's club is based on weather difference whereas Poudel et al. let data speak and therefore used a cluster based approach to identify these convergence clubs. Club convergence is observed because different state level agriculture economies are starting at different points and they converge to a different growth path. This idea of club convergence has been advanced by Galor (1996) and Azariadis and Drazen (1990).

Existing literature provides several reasons for differences in total agricultural productivity across states. These factors are R&D spending (McCunn and Huffman 2000), farm size, and human capital (Liu et al. 2010; Poudel et al. 2011) and R&D spending spillover effect (Liu et al. 2010). To identify what factors affect productivity differences across states, researchers have developed a parametric panel model. An ad hoc model specification in a parametric form may bring biased and inconsistency in the model. Therefore, we proposed a test (Blundell and Duncan 1996; Poudel et al. 2007) should be performed to identify if a variable in the model should in fact be entered in a parametric form. We estimate parametric, semiparametric and nonparametric forms of model and then test for the suitability of these different functional forms using a Hong and White's test.

Methods

To test for the club convergence of regional differences in productivity, we used the following model developed by Sala-i-Martin (1996)

$$\text{Var}(\text{LnTFP}) = \alpha_1 + \alpha_2 t + \varepsilon_t$$

Here, the dependent variable is the across state variance of the natural log of TFP in period t , α_i are parameters and ε is a zero mean random error term. If the coefficient associated with t is negative and significant, it indicates convergence.

Given the fact that previous literature (Poudel et al. 2011 and McCunn and Huffman 2001) have found no absolute convergence, we test for a sigma convergence among states. To identify the club of states having the similar productivity growth, we used a cluster analysis approach. We selected the same number of clubs of states as McCunn and Huffman and identified whether there is convergence among those states.

We use three types of estimation procedures (i.e. parametric, nonparametric, and semiparametric) to estimate relevant coefficients and shape of distribution of private research investment, public research investment, extension investment, farm size, private agricultural spillover, public agricultural research spillover, spillover from other industries to agricultural, farmers educational level, and health care supply level in rural areas on agricultural productivity growth i.e. total factor productivity (TFP) in the U.S. states. Generally, the TFP relations with these variables are studied using a parametric model. However, validity of parametric models have been questioned in the literature because of their ad hoc nature.

Nonparametric/semiparametric specification enables us to estimate shapes of the relationship avoiding an ad hoc choice of parametric functional forms. We use nonparametric and semiparametric functional forms using the method suggested by Racine (2007).

Parametric Estimation of Panel Data Model

Consider a parametric model of the form

$$Y_{it} = X'_{it}\alpha + \mu_i + \vartheta_{it} \quad i = 1, 2, \dots, N, \quad t = 1, 2, \dots, T.$$

Here, Y is the index for the productivity growth i.e. total factor production (TFT) and X_{it} is the matrix of parametric independent variables. In this model, we assume that all the available explanatory variables (private research investment, public research investment, extension investment, farm size, private agricultural spillover, public agricultural research spillover, spillover from other industries to agricultural, farmers' educational level, and health care supply level in rural areas) are parametric and α is a vector of their respective coefficient. μ_i is individual specific effects which makes heterogeneity among the individuals. Since 48 states consider in the study are fixed so we use a fixed effect model rather than a random effect model. This implies μ_i is fixed for all the states. We can estimate fixed effect model by a least square method with dummy variables included for each states. However, Green (2006) suggests using mean difference of each variable which removes heterogeneity. One can then run an OLS model on the transformed data. This method, however removes time invariant independent variables, but the entire explanatory variable in our case are continuous, so we can estimate parametric coefficient competently from fixed effect model which are consistent and efficient.

Nonparametric Estimation of Panel Data Model

Let us consider the following nonparametric panel model in which z_{it} are included as nonparametric variables.

$$Y_{it} = g(z_{it}) + u_{it} \quad i = 1, 2, \dots, N, \quad t = 1, 2, \dots, T.$$

Where $g(\cdot)$ is an unknown smooth function, We assume that data are independent across the i states. \mathbf{z}_{it} is of dimension q . In particular, in pure nonparametric model all the explanatory variables mentioned in above section enter as nonparametric variables and all other variables are scalars. Racine 2007 suggests that the standard approach to introduce individual effect μ_i . Assuming \mathbf{Z}_{it} is strictly exogenous. i.e.

$$E(u_{it} | \mathbf{Z}_{i1}, \mathbf{Z}_{i2}, \dots, \mathbf{Z}_{iT}) = 0$$

for all t . Under this assumption a standard nonparametric estimate $g(\cdot)$, the local constant method is

$$\hat{g}(\mathbf{z}) = \frac{\sum_{j=1}^N \sum_{s=1}^T Y_{js} K_{h,j,s,\mathbf{z}}}{\sum_{j=1}^N \sum_{s=1}^T K_{h,j,s,\mathbf{z}}}$$

Where, $K_{h,j,s,\mathbf{z}} = \prod_{l=1}^q h_l^{-1} k\left(\frac{Z_{j,s,l} - Z_l}{h_l}\right)$ is product kernel function.

For fixed effect model, i.e., states are fixed so we consider a case of one-way error component model with

$$u_{it} = \mu_i + \vartheta_{it}$$

Here μ_i is independently and identically distributed with mean zero and constant variance

$D(0, \sigma_\mu^2)$ and ϑ_{it} has zero mean and finite variance. To estimate nonparametric estimate of

$g(\mathbf{z}_{it})$, we need to choose an appropriate kernel function with correct bandwidth. Many kernel

function are available, However, it is well known that, compared to the selection of the

bandwidth, the choice of the kernel function has only a small impact on the properties of the

resulting estimate (Van and Azomahou, 2007). We use the Gaussian kernel with the following

forms

$$K_h(u) = \frac{1}{\sqrt{2\pi}} \text{Exp} \left[0.5 \left(\frac{u}{h} \right)^2 \right]$$

Where, h is the bandwidth. The least square cross validation method is used to select bandwidth of the kernel function.

Semiparametric Estimation of Panel Data Model

One of the disadvantages of nonparametric regression is that nonparametric regression suffer from the curse of dimensionality. When there are large number of observations and explanatory variables, the tendency is to estimate a semiparametric model. In fact, a semiparametric model is one for which some explanatory variables parametric, while the remaining components have unspecified functional forms. The choice of variables in parametric and semiparametric components comes from economic theory, however a prior assumption of variables does not always hold true, so we need to check whether to incorporate an explanatory variable as a parametric or a nonparametric form in the model. We illustrate two basic approaches for the choice of variable in parametric and semiparametric models. We use Robinson's kernel based partial linear model as a semiparametric model. The partially linear panel data regression model with fixed effect is given

$$Y_{it} = X'_{it}\alpha + g(Z_{it}) + u_{it} \quad i = 1, 2, \dots, N, \quad t = 1, 2, \dots, T, \quad 1$$

Let $\hat{g}_Y(Z_{it})$ be the nonparametric estimator of equation (2). And let $\hat{g}_{X,r}(Z_{it})$ be the nonparametric estimator of r^{th} component of X_{it} , when it is considered as a dependent variable with Z_{it} as an independent variable. Let us define new variable by taking difference form their nonparametric estimate as

$$Y_{it}^* = Y_{it} - \hat{g}_Y(Z_{it})$$

$$X_{it}^* = X_{it} - \hat{g}_X(Z_{it})$$

So that

$$\hat{Y}_{it}^* = (\hat{Y}_{it1}^*, \dots, \hat{Y}_{itT}^*)'$$

$$\hat{X}_{it}^* = (\hat{X}_{it1}^*, \dots, \hat{X}_{itT}^*)'$$

The advantage of this transformation is that it eliminates both nonparametric part and individual effects. Then the original regression equation takes the following forms

$$Y_{it}^* = X_{it}^* \alpha + \varepsilon_{it}$$

Then one can estimate α by using least square estimation procedure. The estimated coefficient are:

$$\hat{\alpha} = \left[\sum_{i=1}^N \hat{X}_{it}^* \hat{X}_{it}^{*'} \right]^{-1} \left[\sum_{i=1}^N \hat{X}_{it}^* \hat{Y}_{it}^{*'} \right]$$

The asymptotic distribution of $\hat{\alpha}$ is given by $\sqrt{N}(\hat{\alpha} - \alpha) \rightarrow N(0, V)$,

Where $V = A^{-1}BA^{-1}$ is a positive definite matrix. Moreover, $\hat{V} = \hat{A}^{-1}\hat{B}\hat{A}^{-1}$ is a consistent

estimator of V, where $\hat{A} = N^{-1} \sum_{i=1}^N \hat{X}_{it}^* \hat{X}_{it}^{*'}$ and $\hat{B} = N^{-1} \sum_{i=1}^N \sum_{t=1}^T \sum_{s=1}^T \varepsilon_{it} \varepsilon_{is} \hat{X}_{it}^* \hat{X}_{it}^{*'}$, and Where

$$\varepsilon = \hat{Y}_{it}^* - \hat{X}_{it}^* \hat{\alpha}$$

Variable Selection

In case of a semiparametric model, it is important to identify which variable should be included as a parametric variable and which should be included as a nonparametric variable. Although, nonparametric variable classification can be accomplished using an established economic theory, these theories sometimes fail to place variables in appropriate categories (either parametric or

nonparametric). For this reason, categorization of variables as either parametric or nonparametric in the semiparametric model must be conducted before the model fitting. One of these methods is developed by Blundell and Duncan (1998). Poudel, Paudel, and Bhattarai (2009) have used this method to identify a semiparametric variable in pollution-income relationship. According to this method, we chose non endogenous variable as a nonparametric variable. Let u be the residual of reduced form of the equation, then significance of coefficient of u in augmented nonparametric regression equation implies the choice of variable as parametric components in semiparametric model.

If we estimate nonparametric regression using a spline function, we can use another approach to select parametric and nonparametric variables. In this approach variables are contrasting the deviance (equivalent to likelihood ratio test) for a model that fits a term nonparametrically with the deviance for an otherwise identical model that fits the term linearly (Fox, 2002; Keele, 2008; Hastie and Tibshirani 1990). The likelihood ratio test for additive nonparametric and semiparametric models takes the usual form:

$$LR = -2(\text{LogLikelihood}_0 - \text{LogLikelihood}_1)$$

Where, LogLikelihood_0 is the log-likelihood for the restricted model and LogLikelihood_1 is the log-likelihood for the unrestricted model, the nonparametric regression model. The test statistics under the null hypothesis follows an approximate chi-square distribution, and the degree of freedom is the difference in the number of parameters across the two models. The deviance for a model is simply -2 times the log likelihood. The resulting tests statistics also follows a chi-square distribution with the same degree of freedom of likelihood ratio test.

Model Specification test

Existing studies have proposed several test statistics to compare the suitability of different functional forms (Hong and White, 1995; Fan and Li, 1996; and Zheng, 1996; Li & Wang, 1998). However, these tests statistics consider models with continuous nonparametric variables. Hsiao et al. (2007) modified these tests statistics compatible with both continuous and categorical variables. We used this test to identify the best functional form (parametric, semiparametric and nonparametric). Assume that parametric model is correctly specified. Then, the null and alternative hypotheses are:

H_0 : Parametric Model

H_1 : Nonparametric/Semiparametric Model

The test statistic purposed by Hsiao, Li and Racine (2007) is

$$J_n = n(h_1 \dots h_q)^{1/2} I_n / \sqrt{\Omega}$$

Where

$$I_n = n^{-2} \sum_i \sum_{i \approx j} u_i u_j K_{h,ij}$$

$$\Omega = \frac{2(h_1 \dots h_q)}{n^2} \sum_i \sum_j u_i^2 u_j^2 K_{h,ij}^2$$

Where u_i is residual from parametric model. J_n is distributed $N(0,1)$ under null hypothesis. J_n test diverges to $+\infty$, if H_0 is false. Thus we reject null hypothesis for large value.

Data

Total factor productivity data for each state between 1960 and 2004 were obtained from the USDA/ERS.

The TFP values are calculated taking Alabama 1996 as the base period. Explanatory variables to describe

the difference in productivity across states are obtained from Liu et al. (2011). Explanatory variables in the model are Private agricultural spillovers, Public agricultural spillovers, Private research investment, Spillovers from other Industries, Education, Health care supply level, Extension investment, Public research investment and Farm size. The details on these variables can be found in Liu et al. (2011). Summary statistics of these variables are provided in Table 5.

Results

Table 1 provides the list of states in each club (clusters). We have identified 10 clubs of states to be consistent with McCunn and Huffman's regional groups. We identified these clusters for the case of 42 states and 48 states. Tables 2,3, and 4 provide results from the sigma convergence test for each of the cluster group for the case of McCunn and Huffman group of states as well as for the club of states we have identified using a cluster analysis method.

In the McCunn and Huffman identified group of clubs, we found that four clusters (1, 2, 4, and 10) had convergence, four clusters had divergence (5, 7, 8, 9) and two had neither convergence nor divergence as the parameter associated with t variable was insignificant. Compare this to statistically derived group of clubs. In 42 states and 10 clubs, we found that clusters (3, 10) of states showed convergence of TFP variance over time. Similarly in 48 states and 10 clubs, we found that clusters 2, 3, and 7 showed convergence in productivity over time. It is clearly evident that the statistical method identified right type of clusters and statistical test showed the convergence among these identified clubs.

For comparisons and interpretation, we also estimated a panel parametric model to identify variables affecting TFP. First, we assumed that each state has fixed effects on the total factor productivity. The F-test statistics for the null of absence of fixed state-specific effect on

agricultural productivity growth was 64.61 which was significant at less than 1% level of significance implying the rejection of a null hypothesis. Therefore, state-specific effects do exist. We also performed Hausman test to examine whether random effect model provided consistent estimators. The test statistics is significant at less than 0.001 suggesting that a random effects model is inconsistent so we should use a fixed effects model. The estimated coefficients for parametric and semiparametric coefficients are provided in Table 3. All three model fit curves for different explanatory variables are shown in Figure 1.

As we discussed in the methods section, we used both Blundell and Duncan's approach and Deviance test to identify parametric and nonparametric variables in semiparametric model. The test results are shown in Table 6. Blundell and Duncan approach test indicated that only public agricultural research spillover is not significant at 5% level of significance which implies that agricultural research spillover is entered as nonparametric component in a semiparametric model. In contrast, likelihood ratio test showed that all the explanatory variables used in this research have nonlinear relationship with TFP. Hence, all the variables are entered as nonparametric components. Based on this test result, we estimated a nonparametric model. The entire nonparametric fitted curves with each explanatory variable are shown in Figure 1.

We conducted Hsiao, Li and Racine (2007) test to compare estimates from parametric, semiparametric and nonparametric model. The test statistics results are provided in table 8. The Jn test statistics value for the comparison of parametric and semiparametric is 10.44 with a significance level less than 0.01, hence we concluded that a semiparametric model is better than a parametric model. Similarly, the test statistics for the comparison between nonparametric and parametric model is 20.24 which is significant at 1% level. Therefore, nonparametric model is a superior parametric model. The output from each model is shown graphically in Figure 1.

Different explanatory variables have different effects on total factor productivity. We found that a semiparametric model is better than parametric model so we interpret coefficients from a semiparametric model. Further we describe nonparametric estimates of each variable graphically.

The estimated coefficient of private agricultural spillovers is positive and significant at 5% level, which implies that increase in private agricultural spillover increases total factor productivity. Similarly, the public agricultural spillovers is also positive and significant in parametric model, however this variable is entered as nonparametric components in a semiparametric model. The graphical estimate in Figure 1(b) shows that public agricultural spillover is also an important factor that increases TFP. In contrast, the estimated coefficient for spillover from other industries is negative and significant. This results shows that other industries has negative influence on TFP.

Private and public research investment also play important roles in factor productivity (McCunn and Huffman, 2000; Liu et al. 2011). Our result suggests public research investment has significant and positive influence on TFP. The nonparametric estimates also match with the result from the semiparametric model as the curve is increasing as shown in Figure in 1. In contrast, semiparametric and nonparametric estimates show that private research investment has negative impact on TFP.

Farmers' education level plays an important role in the agricultural production as educated farmers can use their knowledge to produce commodity efficiently that help to increase agricultural productivity. As expected, our result implies that education has positive impact on total agricultural productivity. Farmers activity on farms depend their health status. We assume that more healthy workers can spend more time and pay more attention on their farm.

Consequently, it helps to produce more output. Our result is consistent with our anticipations the estimated coefficient for health is positive and significant.

Farm sizes capture not only the gross value of product but also the productivity of a farm (Liu et al 2011). The estimated coefficient for farm size is positive and significant with the highest value among all explanatory variables. Figure 1(i) shows that value of TFP in all the three fitted model increases with increase in farm size. Moreover, the nonparametric curve does not goes decline suggesting that agricultural productivity growth always increases with increase in farm size.

Conclusions

Our study indicated that there are only few converging group of states when comes to agriculture total factor productivity. The regional grouping of countries based on weather gave convergence and divergence group of the clubs whereas the statistical method identified only converging group of states.

We used parametric, nonparametric and semiparametric models to explain the role of different variables on productivity differences across 48 states using panel data. Our results show that public research spillover behaves nonparametrically in semiparametric model. The comparison of a parametric model with nonparametric and semiparametric models indicates that nonparametric and semiparametric model is better than parametric model. This paper examines the impact of public and private research spillover and investment on the total factor productivity. We find that private research investment and spillover from other industries have negative impact on the TFP. This may indicate that funds for research are diverted from

agriculture to other alternative uses impacting agriculture negatively. We also examined the impact of health care supply on agricultural productivity growth. The farm size has the highest impact on agricultural production growth.

References

- Azariadis, C., A. Drazen. 1990. "Threshold externalities in economic development." *Quarterly Journal of Economic Development* 105(2):501–526.
- Ball, V.E., C. Hallahan, and R. Nehring. 2004. "Convergence of Productivity: An Analysis of the Catch up Hypothesis within a Panel of States." *American Journal of Agricultural Economics* 86: 1315–21.
- Blundell, R., and A. Duncan. 1998. "Kernel Regression in Empirical Microeconomics." *The Journal of Human Resources* 33(1998):62–87.
- Fan, Y. and Q. Li. 1996. "Consistent Model Specification Tests: Omitted Variables and Semiparametric Functional Forms." *Econometrica*, 64:865-890.
- Galor, O., 1996. "Convergence? Inferences from theoretical models." *Economic Journal* 106 (437):1056–1069.
- Hastie, T.J. and Tibshirani, R.J. 1990. *Generalized Additive Models*, New York: Chapman and Hall.
- Hong, Y. and H. White. 1995. "Consistent Specification Testing via Nonparametric Series Regression." *Econometrica* 63:1133-1995.
- Hsiao, C. and Q. Li and J.S. Racine. 2007. "A Consistent Model Specification Test with Mixed Categorical and Continuous data," *Journal of Econometrics* 140:802-826.
- Keele, L. 2008. *Semiparametric Regression for the Social Sciences*, John Wiley and Sons.
- Li, Q. and S. Wang. 1998. "A Simple Consistent Bootstrap Test for a Parametric Regression Function." *Journal of Econometrics* 87:145-165.
- Li, Q., and J.S. Racine. 2007. *Nonparametric Econometrics: Theory and Practice*, Princeton University Press.
- Liu, Y., C. R. Shumway, R. Rosenman, and V.E. Ball. 2011. "Productivity Growth and Convergence in US Agriculture: New Cointegration Panel Data Results." *Applied Economics* 43(2011):91 –102.
- McCunn, A. and W.E. Huffman. 2000. "Convergence in U.S. Productivity Growth for Agriculture: Implications of Interstate Research Spillovers for Funding Agricultural Research." *American Journal of Agricultural Economics* 82:370-388.
- McCunn, A., and W. Huffman. 2000. "Convergence in U.S. Productivity Growth for Agriculture: Implications of Interstate Research Spillovers for Funding Agricultural Research." *American Journal of Agricultural Economics* 82:370–88.

- Mukherjee, A., and Y. Kuroda. 2003. "Productivity Growth in Indian Agriculture: Is there Evidence of Convergence Across States?" *Agricultural Economics* 29:43–53.
- Nguyen Van, P., and T. Azomahou. 2007. "Nonlinearities and Heterogeneity in Environmental Quality: An Empirical Analysis of Deforestation." *Journal of Development Economics* 84(1):291-309.
- Phillips, P. C. B. and D. Sul. 2007. "Transition Modeling and Econometric Convergence Tests." *Econometrica*, 75: 1771–1855
- Poudel B., K. Paudel, and K. Bhattarai. 2009. "Searching for Environmental Kuznets Curve in Carbon Dioxide Pollutant in Latin American countries." *Journal of Agricultural and Applied Economics*, 41:13–27.
- Poudel, B.N, K. Paudel, and D. Zilberman. 2011. "Convergence in Total Factor Productivity in US Agriculture: Myth or Reality?" *Journal of Agricultural and Applied Economics*, 43(2011):143–156.
- Poudel, B.N., K. P. Paudel, and D. Zilberman. 2011. "Convergence in Agriculture: Myth or Reality?" *Journal of Agricultural and Applied Economics* 43:143-156.
- Rezitis, A.N. 2005. "Agricultural Productivity Convergence across Europe and the United States of America." *Applied Economics Letters* 12: 443–46.
- Rezitis, A.N. 2010. "Agricultural Productivity and Convergence: Europe and the United States." *Applied Economics* 42:1029–44.
- Sala-i-Martin, X. X. 1996. "The Classical Approach to Convergence Analysis," *Economic Journal*, 106:1019–1036.
- Thirtle, C., J. Piesse, A. Lusigi, and K. Suhariyanto. "Multi-Factor Agricultural Productivity, Efficiency and Convergence in Botswana, 1981–1996." *Journal of Development Economics* 71(2003):605–24.
- Zeng, J. X. 1996. "A consistent Test of Functional Form via Nonparametric Estimation Techniques," *Journal of Econometrics* 75: 263-289

Table 1: Cluster of States in the Convergence Check (10 clusters)

Cluster	Based on McCunn and Huffman (2000)	Based on Cluster Analysis	
		42 States	48 States
1	NY, NJ, PA, DE, MD	MT, NM, OK, TN, TX	AR, IA, ID, IL
2	MI, MN, WI	KY, LA, MI, MO, ND, NV, OH, OR, PA, SD, UT, VA, WI	WV, WY
3	OH, IN, IL, IA, MO	AL, AZ, IN, MN, NE, NJ, SC	AL, AZ, IN, MA, MD, NE, NJ, SC
4	ND, SD, NE, KS	CA	CA, FL
5	VA, WV, KY, NC, TN	CO, KS, MD, MS, NY	KS, KY, LA, MO, ND, NH, NV, PA, SD, TN, TX, UT, VA, VT, WI
6	SC, GA, FL, AL	DE	CO, CT, ME, MI, MN, MS, NY, OH, OR
7	MS, AR, LA	FL	DE
8	OK, TX	GA, NC, WA	RI
9	MT, ID, WY, CO, NM, AZ, UT, NV	AR, IA, ID, IL	GA, NC, WA
10	WA, OR, CA	WV, WY	MT, NM, OK

Table 2. Convergence Check Based on 10 Clusters
for 42 states

	Estimates		Standard Error
Cluster1			
Φ_1	-0.21820		0.141610
Φ_2	0.00011		0.000072
Adj. R^2	-0.00580		
Cluster2			
Φ_1	0.03234		0.093820
Φ_2	-0.00001		0.000047
Adj. R^2	-0.02210		
Cluster3			
Φ_1	0.24209	***	0.068820
Φ_2	-0.00012	***	0.000035
Adj. R^2	0.12070		
Cluster5			
Φ_1	0.11176		0.080520
Φ_2	-0.00005		0.000041
Adj. R^2	0.04690		
Cluster8			
Φ_1	-0.12196		0.080510
Φ_2	0.00006		0.000041
Adj. R^2	0.04330		
Cluster9			
Φ_1	0.06515		0.072830
Φ_2	-0.00003		0.000037
Adj. R^2	-0.01520		
Cluster10			
Φ_1	1.99115	***	0.158850
Φ_2	-0.00100	***	0.000080
Adj. R^2	0.60720		

Table 3. Convergence Check Based on 10 Clusters for 48 states

	Estimates	Standard Error
Cluster1		
Φ_1	0.06515	0.072830
Φ_2	-0.00003	0.000037
Adj. R^2	-0.01520	
Cluster2		
Φ_1	1.99115 ***	0.158850
Φ_2	-0.00100 ***	0.000080
Adj. R^2	0.60720	
Cluster3		
Φ_1	0.16145 **	0.076630
Φ_2	-0.00008 **	0.000039
Adj. R^2	0.08860	
Cluster4		
Φ_1	-0.00914	0.071850
Φ_2	0.00001	0.000036
Adj. R^2	-0.02270	
Cluster5		
Φ_1	-0.07641	0.083720
Φ_2	0.00004	0.000042
Adj. R^2	-0.00220	
Cluster7		
Φ_1	0.64472 ***	0.105810
Φ_2	-0.00032 ***	0.000053
Adj. R^2	0.54240	
Cluster9		
Φ_1	-0.12196	0.080510
Φ_2	0.00006	0.000041
Adj. R^2	0.04330	
Cluster10		
Φ_1	-0.17063	0.231630
Φ_2	0.00009	0.000117
Adj. R^2	-0.01780	

Table 4. Convergence Check Based on McCunn and Huffman

	Estimates		Standard Error
Cluster1			
Φ_1	0.38527	*	0.211970
Φ_2	-0.00018	*	0.000107
Adj. R^2	0.02650		
Cluster2			
Φ_1	1.24417	***	0.192670
Φ_2	-0.00062	***	0.000097
Adj. R^2	0.53590		
Cluster3			
Φ_1	0.13926		0.148120
Φ_2	-0.00006		0.000075
Adj. R^2	-0.00950		
Cluster4			
Φ_1	0.63492	**	0.286940
Φ_2	-0.00031	**	0.000145
Adj. R^2	0.05260		
Cluster5			
Φ_1	-2.04282	***	0.281260
Φ_2	0.00107	***	0.000141
Adj. R^2	0.37640		
Cluster6			
Φ_1	-0.00086		0.279710
Φ_2	0.00002		0.000141
Adj. R^2	-0.02280		
Cluster7			
Φ_1	-0.50431		0.307920
Φ_2	0.00026	*	0.000155
Adj. R^2	0.06260		
Cluster8			
Φ_1	-0.40022	***	0.088790
Φ_2	0.00020	***	0.000045
Adj. R^2	0.28250		

Cluster9			
Φ_1	-2.07778	***	0.249760
Φ_2	0.00107	***	0.000126
Adj. R^2	0.62110		
Cluster10			
Φ_1	2.77165	***	0.355690
Φ_2	-0.00136	***	0.000179
Adj. R^2	0.51850		

Table 5. Summary statistics of variables

Variables	Mean	Std. Dev	Minimum	Maximum
Year	1977.5	10.3913	1960	1995
TFP (\$1/100)	74.4180	20.0384	32.7909	162.6036
Private agricultural spillovers	2.4826	4.2242	0.0146	20.8566
Public agricultural spillovers (\$000,000)	77.4574	30.4800	11.0034	146.0720
Private research investment	0.2951	0.4743	0.0025	5.7093
Spillovers from other Industries	158.2515	172.1669	1.9439	584.8289
Education	2.7941	1.7857	1.3971	72.2694
Health care supply level	10.6793	4.2106	4.7571	29.7347
Extension investment (\$000,000)	7.2662	5.6433	0.2540	35.0620
Public research investment (\$000,000)	19.5155	16.8833	0.7508	111.6128
Farm Size (,000)	0.2818	0.2586	0.0139	3.1732

Table 6. Variable Selection

Variable	Blundel & Approach		Devience Method	
	ρ	Variable Type	Devience	Vartiabile type
Private agricultural spillovers	0.0067 (0.032)	Parametric	-4563.312 (0.000)	Nonparametric
Public agricultural spillovers	-0.0001 (0.377)	Nonparametric	-2352.2872 (0.000)	Nonparametric
Private research investment	-0.4013 (0.000)	Parametric	-3574.1678 (0.000)	Nonparametric
Spillovers from other Industries	-4.2079 (0.000)	Parametric	-1045.4639 (0.000)	Nonparametric
Education	-0.1373 (0.000)	Parametric	-564.0259 (0.010)	Nonparametric
Health care supply level	-0.0105 (0.000)	Parametric	-4504.484 (0.000)	Nonparametric
Extension investment	-0.0087 (0.000)	Parametric	-1069.5197 (0.000)	Nonparametric
Public research investment	-0.0029 (0.000)	Parametric	-1989.6391 (0.000)	Nonparametric
Farm Size	-0.0001 (0.000)	Parametric	-2265.0694 (0.000)	Nonparametric

Note: Number in parenthesis are P-value

Significant variable implies they are selected as parametric functional forms in Blundel and Duncan approach

Significant variable implies they are selected as nonparametric functional forms in Devience method

P = coefficient of reduced form equation

Table 7. Parametric and semiparametric model to identify the variables affecting total factor productivity

Variable	Parametric	Semiparametric
Intercept	26.6716 (0.000)	-0.2631 (0.000)
Private agricultural spillovers	0.0817 (0.574)	1.7101 (0.000)
Public agricultural spillovers	0.1053 (0.000)	
Private research investment	9.0412 (0.000)	-7.6487 (0.000)
Spillovers from other Industries	0.0695 (0.000)	-0.0394 (0.000)
Education	0.1527 (0.708)	0.2658 (0.066)
Health care supply level	2.5407 (0.000)	0.9778 (0.000)
Extension investment	0.8934 (0.000)	1.0962 (0.000)
Public research investment	0.2169 (0.000)	0.3722 (0.000)
Farm Size	0.5918 (0.675)	22.5972 (0.000)

Note:

Test for there is no fixed effect is 64.61 with P-value <0.0001

Housman test for random effect vs. fixed test is 16.89 with P-value 0.0047

Table 8. Model Comparisons

Model	Jn
Parametric vs. Nonparametric	20.2466 (0.000)
Parametric vs. Semiparametric	10.4454 (0.000)

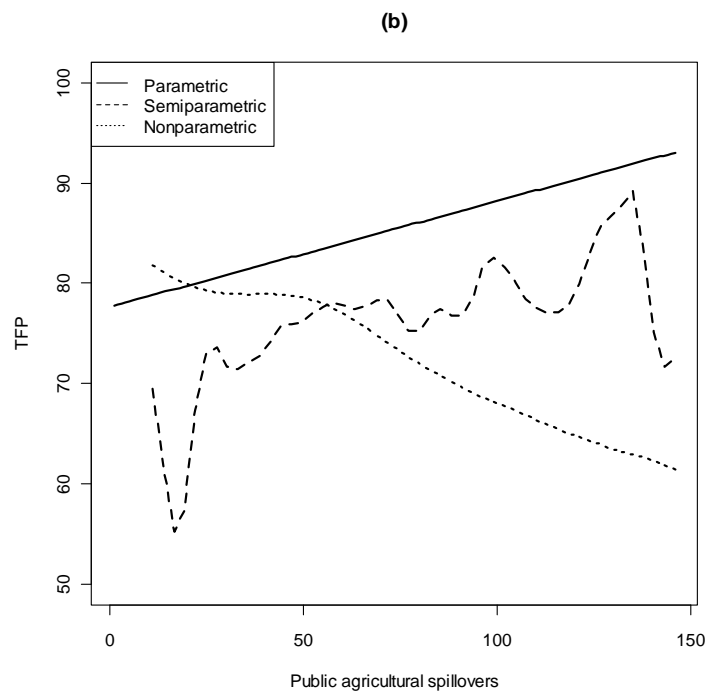
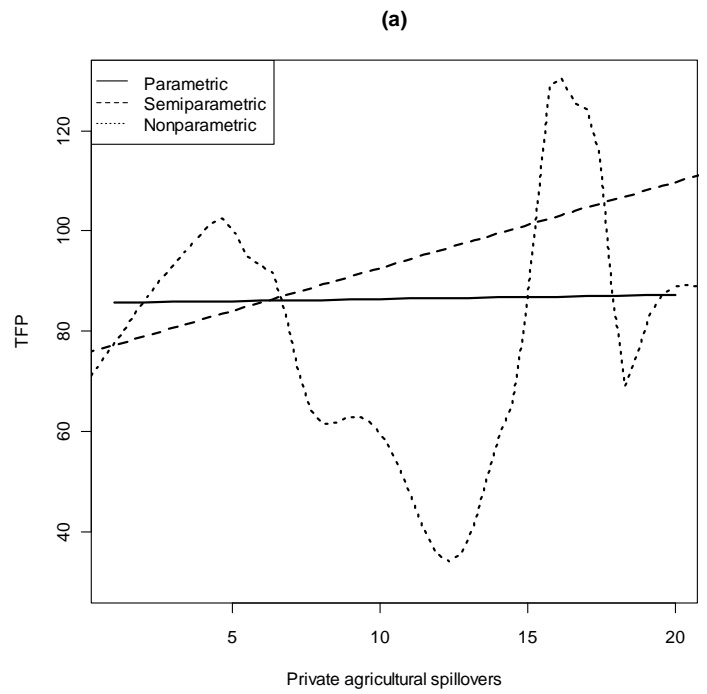


Figure 1 : Parametric, nonparametric, and semiparametric estimation of the relationship between TFP and explanatory variables.

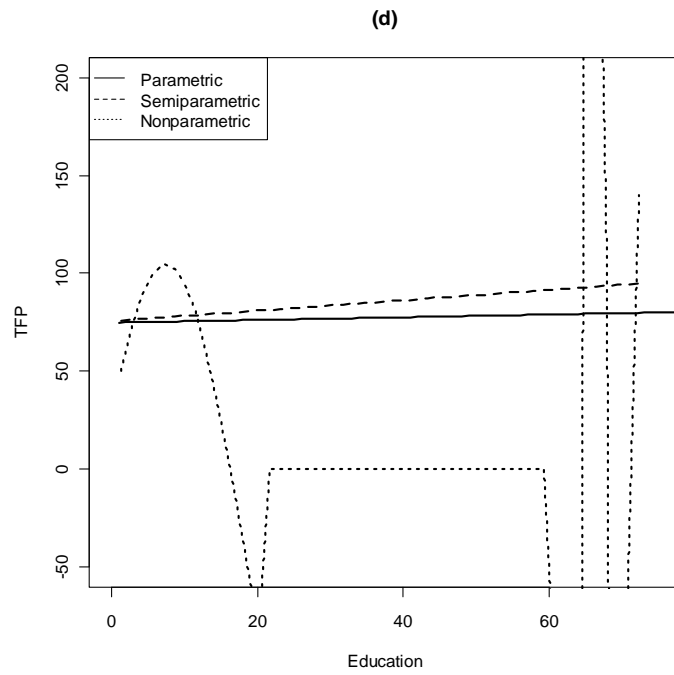
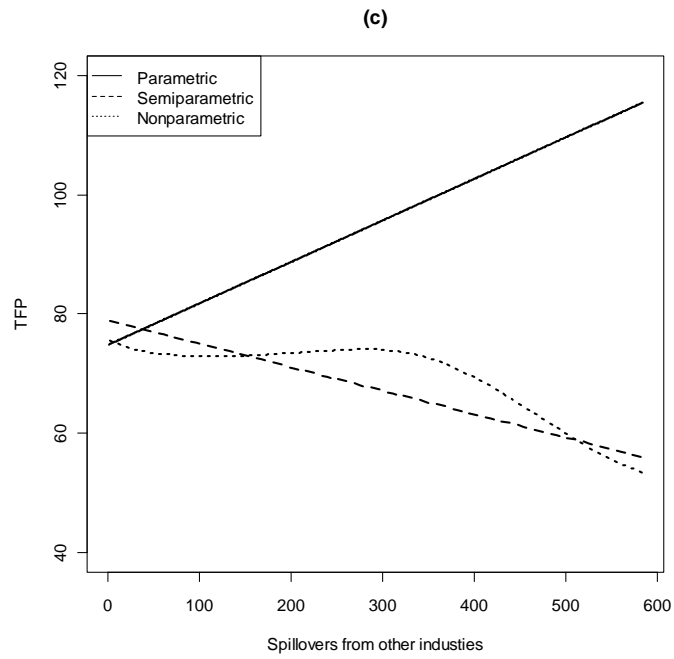
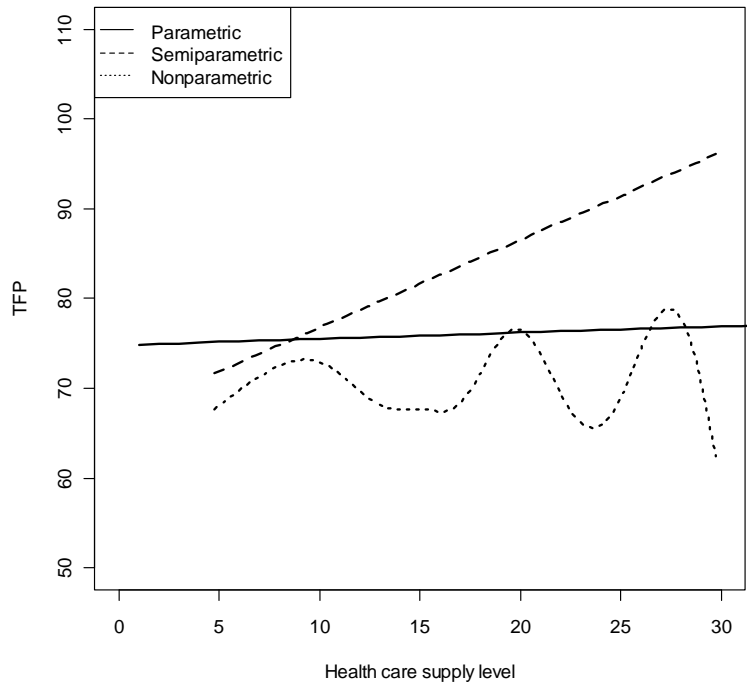


Figure 2 : (Continued.)

(e)



(f)

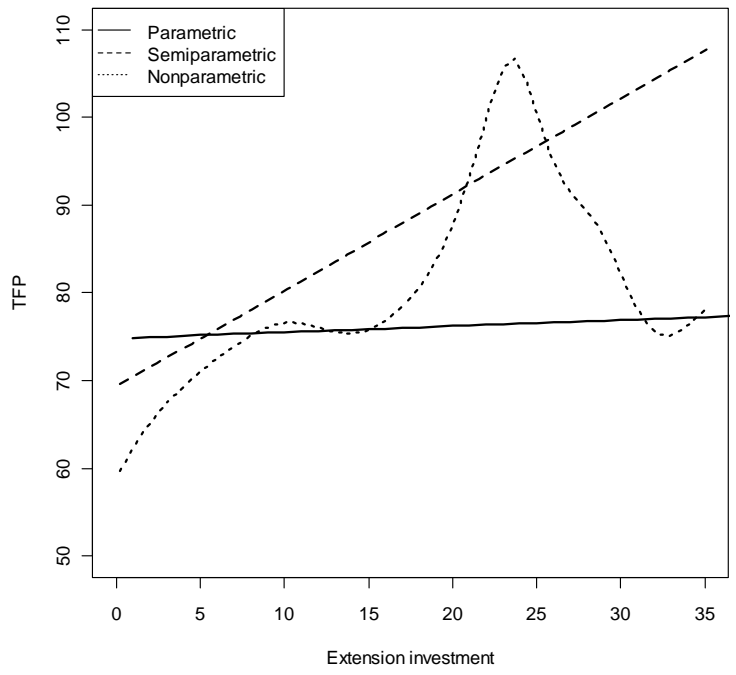
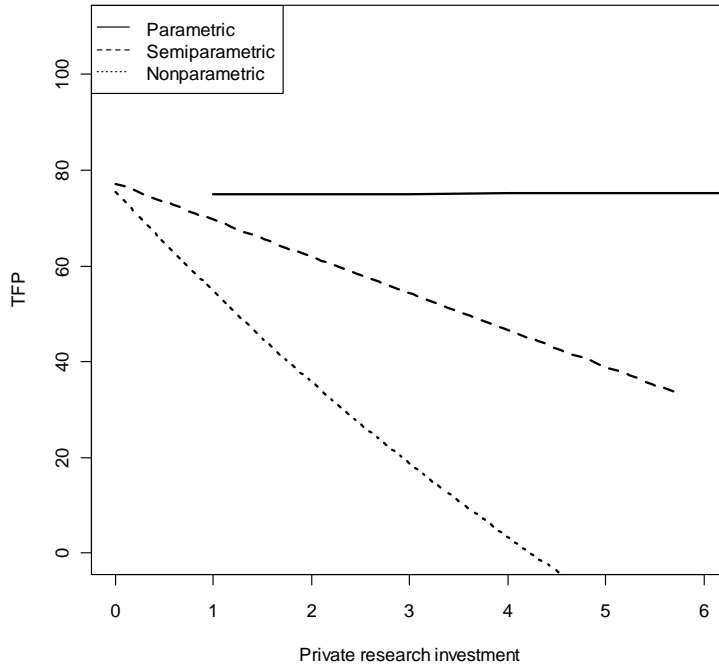


Figure 3 : (Continued.)

(g)



(h)

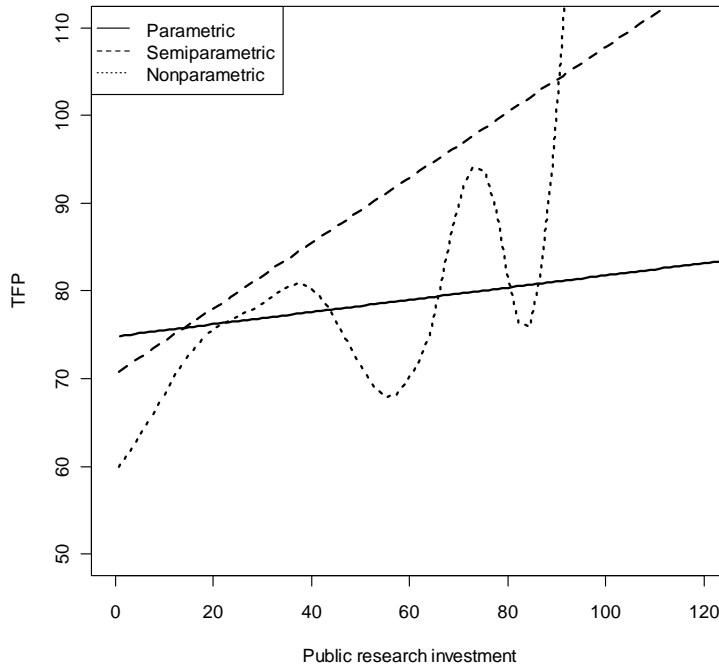


Figure 4 : (Continued.)

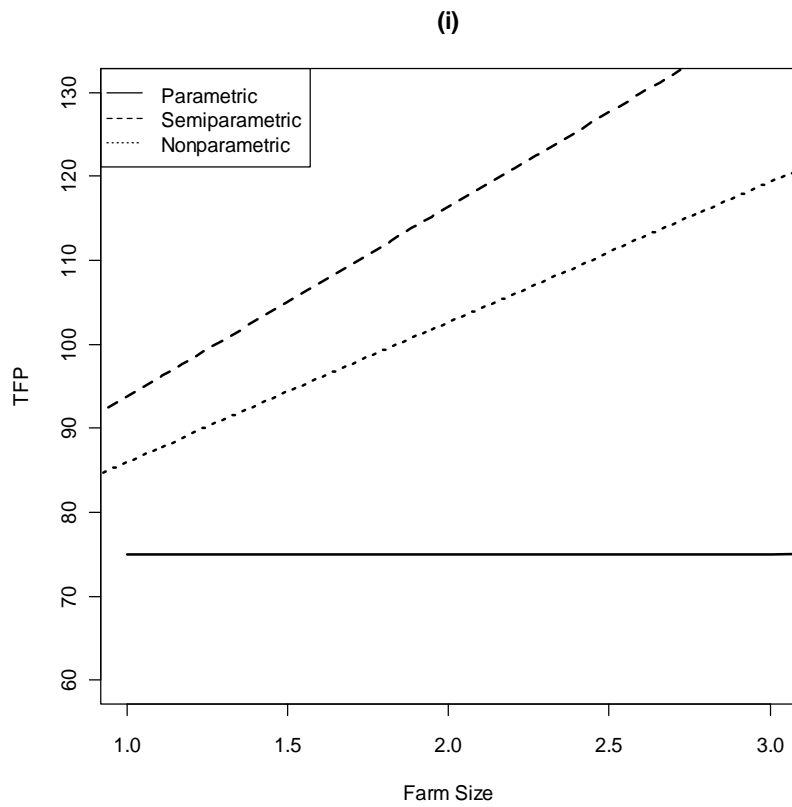


Figure 5 : (Continued.)