No. 940

How Important is Non-farm Sector for Agriculture: A Dynamic Panel Approach for Rural India

by

Anit N. Mukherjee and Yoshimi Kuroda

August 2001

How Important is Non-farm Sector for Agriculture:

A Dynamic Panel Approach for Rural India

by

Anit N. Mukherjee*

Graduate Program, Institute of Policy and Planning Sciences,
University of Tsukuba, Ibaraki, Japan.

and

Yoshimi Kuroda

Institute of Policy and Planning Sciences,
University of Tsukuba, Ibaraki, Japan.

* *Corresponding Author:* Tennodai 1-1-1, Tsukuba Science City, Ibaraki 305-8573, Japan.
Tel+Fax: 81-298-53-9746; E-mail: anit@shako.sk.tsukuba.ac.jp

Date of Submission: August 6, 2001

How Important is Non-farm Sector for Agriculture: A Dynamic Panel Approach for India

Introduction

The non-farm sector in developing countries has come under the spotlight with the increasing realization among policymakers that a comprehensive approach to the alleviation of chronic poverty in rural areas of developing countries must focus on both agriculture and non-agriculture. A viable industrialization strategy in the rural areas needs to understand the characteristics of the rural non-farm sector, and its linkages with the other important rural sector, namely agriculture.

There is general agreement that in the initial stages of the growth process of any developing country, a growing agriculture helps to sustain the rural non-farm economy. The importance of the non-farm sector increases with increasing technological advancement in agriculture, both from the demand and the supply sides (see Lanjouw and Lanjouw, 1995, for an extensive discussion of the literature).

Improvement in productivity in agriculture requires a flow of both consumption and investment goods from the non-farm sector. Although a part of this demand can be met through imports from the urban areas, although such a process would entail various costs (such as transportation, inventory etc.) to the agricultural sector. In such a case, a well-developed non-farm sector in the rural areas supplying such goods and services is efficient both from the production and cost aspects. On the other hand, increasing specialization in agriculture and technological improvement gives rise to opportunities for industries such as food processing and allied activities, that are ideally located in the rural areas close to the source of production of their inputs. In such a case, the value addition that takes place in this process would benefit both agriculture and non-agriculture in the rural areas.

Another important aspect that has come to light in the recent past is the linkage between agriculture and non-agriculture in rural areas through income from labor market participation. It has been seen in previous studies that non-farm sector income provides nearly one-third of the total income for rural households in parts of Africa and Asia (Chuta and Lieldholm, 1978; Lieldholm and Kilby, 1989). A host of studies on poverty, especially using micro-level data for India have indicated that income from off-farm work can be used for consumption and risk-smoothing in years of idiosyncratic agricultural shocks (Walker and Ryan, 1990; Townsend, 1994; Kochar, 1999). A lowering of such risk and smoothing of future consumption would logically lead to greater investment in agriculture, since it is the main occupation of a majority of the population in the rural areas.

Therefore, on the one hand, the development of nonfarm sector generates more income in the rural sector and may lead to greater investment in agriculture; on the other hand, development of the nonfarm sector may also cause an increase in labor's opportunity cost in agricultural production by means of an increase in the employment choice of the rural workers and through its consumption and risk smoothing effect. This may result in lower agricultural labor input. Therefore, the net effect of nonfarm sector development on TFP is an empirical question subject to test and this paper is a step towards filling this gap in the literature.

The main problem faced in a viable econometric analysis of the effect of the non-farm sector on agricultural productivity is the scarcity of a consistent set of data on the former. As is evident from the discussion above, the direction of the causality is also open to question and as yet not addressed specifically by any study. Moreover, the aggregate data in levels of agricultural production and non-farm sector suffer from the familiar problems of endogeneity, common trends, measurement errors etc. A common way to deal with this kind of problem is to use some form of first differencing. However, as the literature on the relation between public capital and productivity has shown, differencing actually destroys the long-run relation and the estimates reflect the short-run effects (See Aschauer, 1989; Garcia-Mila, et.al.,1996; Munnel, 1992; Holtz-Eakin, 1994; Evans and Karras, 1994).

Our objectives in this paper are the following. Using a panel dataset for 14 major states of India from 1973 to 1993, we address the question of causality between non-farm sector development and total factor productivity in agriculture, using the methods recently proposed by Arellano and Bond (1991) and Blundell and Bond (1998). This uses the properties of dynamic models in a panel framework to estimate the causality relation. Next, using the same methodology, we would estimate the effect of the non-farm variable on agricultural productivity using a set of instruments in the dynamic panel model. In extending the previous research on the productivity of public infrastructure to the non-farm sector in the rural areas, we shall see that using a dynamic panel model gives better and more consistent estimates of the impact than using a model with levels or differences only.

In the following section, we outline the methodology that we adopt for the dynamic panel analysis, explaining the choice of instruments and the tests performed. The next section reports the causality test results, followed by an investigation of the productivity impact of the non-farm sector on agriculture. We conclude the paper summarizing the salient features of this kind of analysis and its importance from a policy perspective.

Dynamic Panel Methodology:

The estimation of the impact of non-farm sector on total factor productivity (henceforth TFP) in agriculture has to take into account the bias that would ensue in case there is no causal relationship among the variables. Reverse causality would yield inconsistent estimates of the parameters under the assumptions of ordinary least squares. Most causality tests in sectoral impact studies (such as those of public infrastructure on productivity) have used time-series data (Aschauer,1989; Holtz-Eakin, 1994). However, it is very difficult to eliminate measurement errors and endogeneity problems with a limited number of observations as most time series are.

Initial conditions and previous information also play a vital role both in causality tests and in the model estimation. Lagged values of dependent variable are often used to proxy for past information that is uncorrelated with the error term, so that they can be used as instruments in the estimation process. Thus the dynamic generalized method of moments (GMM) approach will be used in this paper for causality tests and estimation. This method uses all the information contained in previous lags and levels of dependent variables as instruments.

A Simple AR1 Process:

To illustrate, let us start with the assumption that there are *N* cross-sectional units observed over *T* periods. Let *i* index the cross-sectional observations and *t* the time periods. We also take into account an individual effect ς_i for the ith cross-sectional unit.

$$y_{it} = \alpha y_{it-1} + \eta_i + \upsilon_{it} \qquad (1)$$

We will also assume that ς_i and $\tilde{o}_{it}$ are independently and identically distributed across *i* and have the error component structure:

$$E(\varsigma_i) = 0;\ E(\tilde{o}_{it}) = 0;\ E(\tilde{o}_{it}\varsigma_i) = 0 \qquad \text{for } i = 1,\ldots,N \text{ and } t = 2,\ldots,T \qquad (2)$$

$$E(\tilde{o}_{it}\tilde{o}_{is}) = 0 \qquad \text{for } i = 1,\ldots..,N \text{ and } \forall t \neq s \tag{3}$$

In addition, we have the standard assumption concerning the initial conditions y_{i1} :

$$E(y_{i1}\tilde{o}_{it}) = 0 \qquad \text{for } i = 1,\ldots..,N \text{ and } t = 2,\ldots\ldots,T \tag{4}$$

For estimation in first differences, in the absence of any further restrictions on the process of generating the initial conditions, the autoregressive error components model (1) – (4) implies the following j = 0.5 (T-1)(T-2) orthogonality conditions that are linear in á :

$$E(y_{it-s}\Delta\upsilon_{it}) = 0 \qquad \text{for } t = 3,\ldots\ldots,T \text{ and } s \quad 2 \tag{5}$$

where $\Delta\upsilon_{it} = \tilde{o}_{it} - \tilde{o}_{i,t-1}$.

Thus, the instruments available for the system of difference equations is given as:

Equations	*Instruments Available*
$\Delta y_{i3} = \alpha\Delta y_{i2} + \Delta\upsilon_{i3}$	y_{i1}
$\Delta y_{i4} = \alpha\Delta y_{i3} + \Delta\upsilon_{i4}$	y_{i1}, y_{i2}
.	.
.	.
$\Delta y_{it} = \alpha\Delta y_{i(T-1)} + \Delta\upsilon_{iT}$	$y_{i1}, y_{i2}, \ldots\ldots, y_{i(T-2)}$

The available instruments satisfy the moment restriction in (5) and can be compactly written as $E(\boldsymbol{Z}_i'\bar{\upsilon}_{it}) = \boldsymbol{0}$, where $\boldsymbol{Z}_i$ is the (T-2)×j matrix given by:

$$
Z_i = Z_i^D = \begin{pmatrix} y_{i1} & 0 & 0 & \dots & 0 & 0 & \dots & 0 \\ 0 & y_{i1} & y_{i2} & \dots & 0 & 0 & \dots & 0 \\ . & . & . & & . & . & & . \\ . & . & . & & . & . & & . \\ 0 & 0 & 0 & \dots & y_{i1} & y_{i2} & \dots & y_{i(T-2)} \end{pmatrix} \tag{6}
$$

and $\overline{\upsilon}_{it}$ is the (T-2)×j vector $(\Delta\upsilon_{i3}, \Delta\upsilon_{i4}, \dots\dots, \Delta\upsilon_{iT-2})'$.

AR1 Process Including Exogenous Regressor:

In models with explanatory variables, Z_i may consist of submatrices with the above block diagonal form along with one-column instruments. Suppose there exists a predetermined regressor that is correlated with the individual effect ς_i exhibiting the following property:

$$
\begin{aligned}
E(x_{it}\upsilon_{is}) &= 0 \quad \text{for s} \quad \text{t} \\
& 0 \quad \text{otherwise} \\
\text{and} \quad E(x_{it}\eta_i) & 0
\end{aligned}
$$

then the corresponding Z_i^E matrix is given by

$$
Z_i^E = \begin{pmatrix} y_{i1} & x_{i1} & x_{i2} & 0 & 0 & 0 & 0 & 0 & \dots & 0 & \dots & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & y_{i1} & y_{i2} & x_{i1} & x_{i2} & x_{i3} & \dots & 0 & \dots & 0 & 0 & \dots & 0 \\ . & . & . & . & . & . & . & . & & . & . & . & . & & . \\ . & . & . & . & . & . & . & . & & . & . & . & . & & . \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \dots & y_{i1} & \dots & y_{i(T-2)} & x_{i1} & \dots & x_{i(T-1)} \end{pmatrix}
$$

As stated in Arellano and Bond (1998), where the number of columns in Z_i^E are very large, using the whole history of the instruments in later cross-sections may lead to overfitting bias in small sample empirical study. As we shall see later, we would fix a maximum and minimum lag on the instrument set of the dependent variable and the regressor in our analysis. We shall also undertake the Sargan test of overidentifying assumptions to check whether our instrument set is valid or not.[1]

Combining Differences and Levels:

To get the full set of instruments combining the levels and differences, we note that the error term in our panel data model consists of $\tilde{o}_{it}$ and ς_i. Where there are instruments available that are uncorrelated with the individual effects ς_i, we can use these variables as instruments for the equation in levels (Arellano and Bond, 1998). Under the assumption of mean-stationarity of the model in (1), since Δy_{it} will be uncorrelated with ς_i, $\Delta y_{i(T-1)}$ can be used as instruments in the levels equations. In this case, the instrument matrix can be stated as:

$$Z_i^S = \begin{pmatrix} Z_i^D & 0 & \dots & 0 \\ 0 & \Delta y_{i2} & \dots & 0 \\ . & . & & . \\ 0 & 0 & \dots & \Delta y_{i(T-1)} \end{pmatrix} \tag{7}$$

which is a block diagonal matrix combining the elements of both (6) and the instruments for the level equations in their diagonal elements.

In the case that exogenous regressors are also present in the model, the optimal instrument matrix can be set up combining the matrices for the regressors and the lagged dependent variables.

One-step estimations using a known weighting matrix are efficient when the error term $\tilde{o}_{it}$ is known to be homoskedastic. In the case of heteroskedastic $\tilde{o}_{it}$, two-step estimators that use the variance matrix from the estimated error terms in the first step as

[1] For a full explanation of the estimation process using dynamic panel data, refer to Blundell and Bond (1998) and Arellano and Bond (1998).

weights in the regression perform better. However, as stated in Arellano and Bond (1998), for hypothesis testing purposes, it is better to use standard errors from the first step while using the two-step weighting matrix. Since the errors in our case are likely to be heteroskedastic, we would report the robust one-step estimates unless otherwise stated.

In the following section, we explain the data used for the analysis in detail and present the result of the causality test using the techniques described above.

Tests of Causality

Data

The dataset employed is a panel of fourteen states of India that are considered to be the most economically important from 1973 to 1993. This dataset has been compiled by the World Bank and the International Food Policy Research Institute (IFPRI) in collaboration with various agencies of the Government of India. [2,3]

Productivity growth in agriculture is measured as the TFP index which is the ratio of total output to total input. The Törnqvist-Theil index is used to construct the TFP growth as follows:

$$\ln TFP_t = \sum_i 0.5*(S_{i,t} + S_{i,t-1})*\ln(Y_{i,t}/Y_{i,t-1}) - \sum_i 0.5*(W_{i,t} + W_{i,t-1})*\ln(X_{i,t}/X_{i,t-1}) \quad (8)$$

where $\ln TFP$ is the log of the total factor productivity index; $S_{i,t}$ and $S_{i,t-1}$ are output i's share in total production value at time t and $t-1$, respectively; and $Y_{i,t}$ and $Y_{i,t-1}$ are

[2] The states in alphabetical order are: Andhra Pradesh, Bihar, Gujrat, Haryana, Karnataka, Kerala, Madhya Pradesh, Maharashtra, Orissa, Punjab, Rajasthan, Tamil Nadu, Uttar Pradesh and West Bengal.

[3] A set of tables containing the data for the variables is available from S.Fan, P.Hazell and S. Thorat (1999), 'Linkages between Government Spending, Growth and Poverty in Rural India', Research Report # 110, International Food Policy Research Institute, Washington, D.C.

quantities of output i at time t and $t\text{-}1$, respectively. Farm prices are used to calculate the weights of each crop in the value of total production. $W_{i,t}$ and $W_{i,t-1}$ are cost shares of input i in total cost at time t and $t\text{-}1$, respectively; and $X_{i,t}$ and $X_{i,t-1}$ are quantities of input i at time t and $t\text{-}1$, respectively. Thirty crops (rice, wheat, jowar, bajra, maize, ragi, barley, gram, other pulses, groundnut, sesame, linseed, rapeseeds and mustard, castorseed, safflower, nigerseed, coconut, soybeans, sunflower, potato, tapioca, sweet potato, banana, cashewnut, coffee, jute, sugarcane, onion and fruits) and three major livestock products (milk, chicken, and sheep and goat meat) are included in total production. Farm prices are used to calculate the output shares.

Five inputs (labor, land, fertilizer, tractors and animals) are included. Labor input is measured as total female and male labor (including both family and hired) engaged in agricultural production. A conversion ratio of 0.7 has been used to convert female labor to its male labor equivalent. Land is measured as net cropped area; fertilizer input is measured as the total amount of nitrogen, phosphate and potassium uses; tractor input is measured by the number of four-wheel tractors (including both private- and government-owned); and animal input is measured as the number of draft animals (total buffalos). Wages of agricultural labor are used as the price of labor; rental rates of tractors and animals are used for their respective prices; and fertilizer price is calculated as a weighted average of the prices of nitrogen, phosphate and potassium. The land price is measured as the residual of total revenue net of measured costs for labor, fertilizer, tractors and bullocks.

NAG is the total number of persons engaged in non-agricultural activities in the rural areas. Employment is defined on the basis of usual status (more than 50 percent of time) of workers in a particular employment category. The data is collected by the National Sample Survey Organization (NSSO) every five years beginning in 1973.

There are two kinds of surveys that the NSSO carries out under various heads. The quinquennial data that we use are the 'full-sample' years that are more reliable as compared to the intervening 'thin-sample' years. We convert the quinquennial data to an annual series taking the population growth rates for the respective states as weights. However, estimations using only the data for the available series have not shown substantial differences in estimates in this and earlier studies.

The results of the causality test are presented in Table 1. We include six lags of both the TFP and the NAG variables for estimation. Although in most of the analysis of standard time-series models lag lengths of less than four are considered, ideally we should test for causality using an arbitrarily long lag length. However, as noted by Holtz-Eakin et.al (1998), the optimal lag-length should be less than one-third of the total time period to avoid overidentification problems.

The NAG variable is significant in the third and fourth lag for the TFP equation in Table 1. For the NAG equation, only the fifth lag of the TFP variable is significant. We perform a Wald test under the null hypothesis of all coefficients of the explanatory variables are jointly zero in each equation. We can see that in both cases, the hypothesis is rejected at one percent level, indicating the existence of bi-directional causality in the two variables. Moreover, we do not find any evidence of first-order serial correlation in the error term in both cases. We can thus use the techniques described in the previous section to efficiently estimate the productivity impact of nonfarm sector on agriculture using an appropriate econometric model, as we shall do in the next section.

Productivity Impact of the Non-farm Sector:

The two-way causality tests have shown that there may exist endogeniety problems associated with the determination of the TFP and the non-farm employment in India. This may be due to regional variation in the development of the non-farm sector, with some regions having higher levels of such employment than others. Using standard ordinary least squares techniques will lead to serious errors in estimation. Although we can get estimates that are robust for heteroskedasticity and reduce the spatial effect by using regional dummies, it can still be subjected to the criticism of endogeneity. The problem then is to find the appropriate instruments that are uncorrelated not only with the error term but also with the regional effect. Therefore we use the dynamic panel technique we have explained in Section II.

We would undertake the estimation under three different model specifications with all the variables in their natural logarithms. The first model uses the current non-farm variable as the regressor along with a constant, with regional dummies:

$$\ln TFP_{it} = \alpha_0 + \beta_1 \ln NAG_{it} + \eta_i + \upsilon_{it} \tag{8}$$

where α_0 denotes the constant term that would capture the common shocks across states in each year and NAG is the non-agricultural employment variable in the model.

As we have seen from Table 1, it might be the case that the TFP of the previous year may have a substantial effect on the current value of the variable. In reality, this is more likely to be the case since the cultivators base their decisions on historical levels of the productivity while choosing their optimal production inputs. Thus we have to consider an auto-regressive model as described in Section II in order to control for the effect of the past productivity level.

$$\ln TFP_{it} = \alpha_0 + \beta_1 \ln TFP1_{it} + \beta_2 \ln NAG_{it} + \eta_i + \upsilon_{it} \tag{9}$$

In this case as well, the ordinary least squares estimate would be upwardly biased in the case of the TFP variable since the estimation would take into account only the first lag of the variable, whereas the system estimator would include further lags. The information contained in the additional instruments would better reflect the optimal decision-making process of the cultivators and we would expect the coefficients to be more efficient.

In the same vein, the non-farm sector employment would also depend on past values and should be included in the estimation.

$$\ln TFP_{it} = \alpha_0 + \beta_1 \ln TFP1_{it} + \beta_2 \ln NAG_{it} + \beta_3 \ln NAG1_{it} + \eta_i + \upsilon_{it} \qquad (10)$$

We would expect the coefficient on the current value to be positive, because a higher employment in the non-farm sector would imply better supply of non-farm goods to the rural sector, that in turn would mean a higher level of productivity for the rural sector as a whole. If the total effect $\beta_1 + \beta_2$ is positive, then we can infer that both the lagged and the current values of the non-farm sector have a significant effect on productivity in the rural areas.

In each of the above cases, we shall compare the levels, differences and the dynamic panel formulation for estimation. The level and difference formulations use the ordinary least squares corrected for heteroskedasticity and fixed effect dummies, which are regional dummies for the former and time dummies for the latter. The system method is estimated with a constant and the regional dummies to control for heterogeneity across states.

Following Arellano and Bond (1998), we restrict the number of instrument that we use for later equations in the model. We use a maximum of six lags for the lagged dependent variable and the regressors to avoid overfitting bias. To test for the presence of bias in the selection of the model due to the use of more instruments than are required, we perform a Sargan test under the null hypothesis of the absence of excess instruments. We also perform stationarity test for the error term under the null hypothesis of presence of first order serial correlation. In case we find significant negative values for the test statistic, we can say with a certain degree of confidence that the estimates are free from biases due to non-stationarity of the data series used. This condition is very strict in the case of the system GMM estimator.

In some cases, the t-statistic may not truly reflect the joint significance of the estimates under scrutiny. We perform the Wald test under the assumption that all the coefficients of the estimates are jointly zero. Even in the case that a particular estimate is not significant, it may so happen that the Wald test may reject the null for joint significance. In such a case, it would still be advisable to include the regressors in the model specification.

The estimation results are given in tables 2, 3 and 4. All the results shown are the one-step estimates with robust test analyzed using the GAUSS program code, *DPD98 for Gauss*, written by Arellano and Bond.

An analysis of the Tables 2 to 4 indicates that our using lagged values of dependent (and exogenous) variables lead to better estimates for the data. Among the three models considered, only the system GMM method gives consistent results for both the serial correlation and the Sargan tests, as well as more dependable estimates of the coefficients. We find that the coefficients for lagged lnTFP and lnNAG are positive and significant in the level and the system specifications. In the difference model, the coefficient for the lagged lnTFP is negative and that for lnNAG is positive but not significant in any model specification. This is consistent with the finding of the literature on productivity of public infrastructure in the U.S (Holtz-Eakin 1994, Evans and Karras, 1994) and a more recent study on India by Zhang and Fan (2001). The reason for this is that in the difference model, only the very short-run effects are evident. As we have seen from the causality test results in Table 1, the productivity effect of the NAG variable is expected over a longer period of time, that the level and the system models try to capture.

Concentrating our attention on Tables 3 and 4, we can infer that there is no random variation in the coefficients for the lagged lnTFP and the lnNAG variables across the three models (the estimates of the lagged lnTFP variable is in fact nearly the same). The estimated effect of the lagged lnTFP variable is negative when estimated in differences, but positive in the other two. The estimated effect is less in the system specification for the lagged lnTFP and more for the lnNAG variable, primarily because in the system specification, a considerable portion of the information contained in the initial conditions and the later instruments is used in the estimation. The test for the presence of serial correlation is rejected at the 10 percent level for the second formulation and at less than 5 percent level for the system method of estimation. This is important since we can then say with a certain degree of confidence that the two series are mean-stationary and the causality test results as well as the estimates are free from the bias of correlation among the error terms. This means that we can use the full set of moments in the estimation as outlined in Section II.

Comparing our results with that of Zhang and Fan (2001), we find that in the regression equation (9), our estimate of the effect on TFP of the non-farm variable (0.363) is substantially higher than the one obtained for road (0.012) for rural India. This can be due to the possibility of a more direct linkage between the non-farm sector and agricultural productivity than the effect of road infrastructure. However, they both point to the need for increasing the focus on rural infrastructure and off-farm employment opportunities in the rural areas.

Including a lagged value of the exogenous regressor in the regression (10) increases the value of the coefficient of lnNAG (Table 4). The coefficient of lnNAG1 is not significant as per the t-statistic. However, the Wald test of joint significance of the two variables rejects the null hypothesis that they are jointly zero. Thus, the absolute effect of the NAG variable is nearly the same as the one obtained by only including the NAG variable in the estimation equation (9). Thus, based on this formulation, we can infer that the effect of the non-farm employment is unambiguously positive and significant for the fourteen Indian states under consideration.

Conclusion

In this paper, we set out to analyze the impact of non-farm sector development on agricultural productivity. Using a panel dataset for fourteen major states of India from 1973 to 1993, we conduct a causality test to determine, firstly, whether the two series of data for the variables are causally related or not, and secondly, to determine the direction of the same. In doing so, we use a set of instruments that are uncorrelated with the errors, using the dynamic panel approach of Arellano and Bond (1991) and Blundell and Bond (1999). As expected, we find that non-farm employment and productivity affect each other in the long run, allowing for individual effects across the states under consideration.

To evaluate the magnitude of the productivity impact of the non-farm sector, estimation procedure is carried out under three separate formulations. In the first formulation only the exogenous variable, NAG, is included. The lagged dependent variable (TFP) and NAG are used as regressors in the next one. The third formulation uses lagged value of the exogenous variable along with the previous variables in the estimation. Using the differenced, level or the system model, we compare the magnitude of the impact under the above formulations. The first formulation and the differenced model do not offer any insight because the effect of previous TFP levels and the long-run relationship is clearly ignored in these cases. However, comparing the levels and the system models, we can infer that the estimates are robust to the alternative formulations and that the impact of the non-farm employment on productivity in agriculture is unambiguously positive and significant, and the magnitude of the effect is of a surprisingly high order.

There are several lessons to be learnt from the analysis. Firstly, contrary to the prevailing wisdom, the non-farm sector not only provides a market for the agricultural sector goods, it has an important role to play in the improvement of productivity in agriculture itself. It is important to recognize that much of the impact of non-farm sector on agricultural productivity has taken place in areas where the level of agricultural productivity is higher, as is evident from the significance of the lagged values of TFP in our analysis. However, the need for non-farm sector development is much greater in areas of depressed productivity growth in agriculture (the so-called 'less favoured areas') since off-farm work can smooth out income and consumption fluctuations and encourage farmers to take risks and invest more in productive inputs in the long run. Our causality test results bear testimony to this aspect of the farm-nonfarm linkage.

Since the non-farm employment in rural areas are primarily in small and medium-scale enterprises, there is an important role of the government in encouraging entrepreneurial activities and create a conducive environment for investment. This study makes an important contribution in clarifying the existence of the linkage between agriculture and non-agriculture sectors in rural areas from a methodological point of view. From a technical perspective, it utilizes the panel data techniques that have been shown to be more powerful than traditional time series methods and extends the debate of rural infrastructure and productivity to the rural non-farm sector in developing countries.

References:

Arellano, M. and S. Bond. "Some Tests of Specification For Panel Data: Monte Carlo Evidence and an Application to Employment Equations." *Rev. Econ. Studies* 58 (1991): 277-297.

Arellano, M. and S. Bond. *Dynamic Panel Data Estimation Using DPD98 for Gauss: A User's Guide*. Memo. (1998).

Aschauer, D. "Is Public Expenditure Productive?" *J. Mon. Econ*. 23 (1989):177-220.

Blundell, R. and S. Bond. "Initial Conditions and Moment Restrictions in Dynamic Panel Data Models." *J. Econometrics* 87 (1998):115-143.

Blundell, R. and S. Bond. *GMM Estimation with Persistent Panel Data: An Application to Production Functions.* Working Paper Series No. W99/4, Institute of Fiscal Studies, London, 1998.

Chuta, E. and C. Lieldholm. *Rural Non-Farm Employment: A Review of the State of the Art*. MSU Rural Development Paper No.4, Michigan State University, East Lansing, 1979.

Evans, P. and G. Karras. "Are Government Activities Productive? Evidence from a Panel of U.S. States." *Rev. Econ. Stat*. 76 (February 1994):1-12.

Fan, S., P. Hazell, and S. Thorat. *Linkages between Government Spending, Growth and Poverty in Rural India.* Research Report 110, International Food Policy Research Institute, Washington, DC, 1998.

Fan, S., P. Hazell and S. Thorat. Government Spending, Growth and Poverty in Rural India, *Am. J. Agr. Econ*. 82 (4) (November 2000):1038-1051.

Fan,S., P.Hazell and T. Hoque. "Targeting public investments by agroecological zone to achieve growth and poverty alleviation goals in rural India." *Food Policy* 25 (2000): 411-428.

Garcia-Mila, T., T. J. McGuire and R.H. Porter. "The Effect of Public Capital in State-level Production Functions Reconsidered." *Rev. Econ. Stat.* 78 (February 1996):177-180.

Holtz-Eakin, D. "Public-Sector Capital and the Productivity Puzzle." *Rev. Econ. Stat.* 76 (February 1994), 12-21.

Kochar, A. "Smoothing Consumption by Smoothing Income: Hours-of-Work Responses to Idiosyncratic Agricultural Shocks in Rural India." *Rev. Econ. Stat.* 81 (February 1999): 50-61.

Lanjouw, J.O., and P. Lanjouw. *Rural Nonfarm Employment: A Survey.* Policy Research Working Paper 1463, World Bank, Washington, D.C., 1995.

Liedholm, C. and P. Kilby. "Nonfarm Activities in the Rural Economy", Eds. J. Williamson and V. Panchamukhi, *The Balance between Industry and Agriculture*. New York: St.Martin's Press, 1989.

Munnel, A. H. "Infrastructure Investment and Economic Growth." *J. Econ. Perspectives* 6 (1992):189-198.

Townsend, R.M. "Risk and Insurance in Village India." *Econometrica* 62 (1994):539-591.

Walker, T.S. and J.G.Ryan. *Village and Household Economies in India's Semi-Arid Tropics*. Baltimore: The Johns Hopkins University Press, 1990.

Zhang, X. and S. Fan. *How Productive is Infrastructure? New Approach and Evidence from Rural India.* Mimeo, International Food Policy Research Institute, Washington, D.C., 2001.

Table 1: Causality Tests for TFP and NAG

	Dependent Variable: TFP		Dependent Variable: NAG	
Variable	Coefficient	P-value	Coefficient	P-value
Const.	-0.344	0.511	0.081	0.448
TFP1	0.311	0.000	-0.028	0.232
TFP2	0.291	0.000	0.014	0.441
TFP3	0.124	0.003	0.028	0.124
TFP4	-0.042	0.649	0.014	0.175
TFP5	0.193	0.000	-0.037	0.002
TFP6	0.209	0.000	0.005	0.659
NAG1	0.799	0.251	1.449	0.000
NAG2	-0.629	0.499	-0.391	0.000
NAG3	-1.236	0.011	-0.013	0.780
NAG4	2.131	0.042	-0.373	0.000
NAG5	-1.439	0.263	0.168	0.211
NAG6	0.368	0.698	0.154	0.031
Wald Test	NAG1-NAG6	0.000	TFP1-TFP6	0.002
Test for Serial Correlation	-2.159	0.031	-2.391	0.017

Table 2: Estimation of $TFP_{it} = \alpha_0 + \beta_{it} NAG_{it} + \eta_i + \upsilon_{it}$

	(1) *Difference (OLS)*	*(2)* *Level (OLS)*	*(3)* *System (GMM)*
Const.	-0.070 (0.053)**	1.757 (0.000)*	0.911 (0.372)
NAG	0.035 (0.805)	0.343 (0.000)*	0.478 (0.000)*
Serial Corr. Test	-2.27 (0.026)**	1.933 (0.053)	-2.236 (0.025)**
Sargan Test			13.18

Note: The figures in parenthesis indicate p-values
*denotes significance at 1% level, **denotes significance at 5% level.

Table 3: Estimation of $TFP_{it} = \alpha_0 + \beta_1 TFP1_{it} + \beta_2 NAG_{it} + \eta_i + \upsilon_{it}$

	(1) *Difference (OLS)*	*(2)* *Level (OLS)*	*(3)* *System (GMM)*
Const.	0.028 (0.406)	0.684 (0.251)	0.348 (0.718)
TFP1	-0.538 (0.000)*	0.504 (0.000)*	0.321 (0.034)**
NAG	0.033 (0.805)	0.191 (0.000)*	0.363 (0.000)*
Serial Corr. Test	0.251 (0.802)	-1.923 (0.055)	-2.053 (0.040)**
Sargan Test			11.39

Note: The figures in parenthesis indicate p-values
*denotes significance at 1% level, **denotes significance at 5% level.

Table 4: Estimation of $TFP_{it} = \alpha_0 + \beta_1 TFP1_{it} + \beta_2 NAG_{it} + \beta_3 NAG1_{it} + \eta_i + \upsilon_{it}$

	(1) *Difference (OLS)*	*(2)* *Level (OLS)*	*(3)* *System (GMM)*
Const.	0.029 (0.387)	0.670 (0.503)	0.024 (0.782)
TFP1	-0.538 (0.000)*	0.505 (0.000)*	0.419 (0.001)*
NAG	0.066 (0.836)	0.241 (0.318)	0.674 (0.046)**
NAG1	-0.044 (0.889)	-0.049 (0.877)	-0.358 (0.293)
Wald Test $H_0: \beta_2 + \beta_3 = 0$	0.043 (0.979)	6.718 (0.035)**	18.44 (0.000)*
Serial Corr. Test	0.274 (0.784)	1.931 (0.054)	-2.087 (0.037)**
Sargan Test			11.61

Note: The figures in parenthesis indicate p-values
*denotes significance at 1% level, **denotes significance at 5% level.