Explaining Production Inefficiency in China's Agriculture using Data Envelope Analysis and Semi-Parametric Bootstrapping

Daniel C. Monchuk†
Department of Economics
University of Southern Mississippi

Zhuo Chen
The Chicago Center of Excellence in Health Promotion Economics
The University of Chicago

Paper to be presented at
The American Agricultural Economics Association Annual Meetings
Orlando, Florida
June 27-29, 2008

Abstract:
In this paper we examine more closely the factors associated with production inefficiency in China's agriculture. The approach we take involves a two-stage process where output efficiency scores are first estimated using data envelope analysis (DEA), and then in the second stage, variation in the resulting efficiency scores are explained using a truncated regression model with inference based on a semi-parametric bootstrap routine. Among the results we find a heavy industrial presence is associated with reduced agricultural production efficiency and may be an indication that externalities from the industrial process, like air and ground water pollution, affect agricultural production. We also find evidence that counties with a large percentage of the rural labor force engaged in agriculture tend to be less efficient, which suggests that policies to facilitate the removal of labor from agriculture, but not necessarily from the rural areas, would bring about enhanced agricultural efficiency and calls into question policies that promote wholesale migration from rural areas. Sensitivity analysis indicates results are robust to influential observations and outliers.

JEL codes: C14, Q1, R5
Keywords: China's agriculture, DEA, bootstrapping, technical efficiency

† Department of Economics, #5072 – 118 College Drive
University of Southern Mississippi
Hattiesburg, MS 39406
Office phone: +1(601)266-6094
email: dmonchuk@gmail.com

I. Introduction:

Being the most populated country and occupying a landmass on par with the U.S. or Europe, China has demonstrated its determination in food self-sufficiency with limited cropland (75 percent that of the U.S.). However, recent World Trade Organization statistics showed that China had a modest $2.9 billion of deficit in agricultural trade in 2000 (Gale 2002). How well China can feed its population in the future remains a question of debate. Nonetheless, further growth in agricultural trade and/or domestic production is expected to accompany the increasing demand of meat poultry, fish, fresh fruit and vegetables, and other high value products of the emerging Chinese middle class (Gale 2002). Given limited per capita resources, if they intend to be competitive in the global market, Chinese agricultural producers need to be efficient. To facilitate continued growth in a sustainable manner, policy to promote greater efficiency of agricultural production will become increasingly important in a bid to limit the strain on the nation's environment, water resources, and infrastructure. Indeed, given the trend of decline in the area of agricultural land (Yao and Liu 1998), technological improvements that shift the production frontier outwards and improvements in technical efficiency are required to increase agricultural production. While the standard response to improve technology has been to increase research and development expenditures, there is no consensus on policies to remedy the problem of technically inefficient production.

The existing literature has generated a considerable discussion on whether agricultural production inefficiencies could be further reduced in China. The role of technical and allocative efficiency was investigated in Chen and Huffman (2006), Mao and Koo (1997), Wang, Cramer, and Wailes (1996), and many others. Abdulai and Huffman (2000) summarized results of several agricultural efficiency studies and found a higher average profit inefficiency score for China (Wang, Wailes and Cramer, 1996) than those of other countries. Tian and Wan (2000) estimated national average efficiency scores ranging from 0.85 to 0.95 for several grain crops and cautioned that quite limited output growth could be attained through input injection and efficiency gains. Following the economic reforms in 1978 some research has argued that although there has been technological progress, at the same time there has also been a decline in technical efficiency due to unfamiliarity of the new technologies (Mao and Koo 1997). If significant inefficiency exists then further increases in output could be achieved through policy that facilitates efficient input and output market operations and creates the institutional and operational framework necessary to make more efficient use of existing inputs. For example, Dong and Putterman (1997) show that better access to credit can increase productivity in Chinese agriculture and Chavas et al. (2005) reach a similar conclusion regarding improving financial markets in The Gambia. Chen, Huffman and Rozelle (2006) applied a stochastic frontier model with farm fixed-effects using a panel of rural households in China and found that mechanization, specialization, and the highest education achieved by household members improve technical efficiency while land fragmentation reduces it. Wang, Wailes, and Cramer (1996) claimed that being a large farm is associated with positive and statistically significant profit efficiency gains.

Many of the above studies have used micro-level data sets. Estimates derived by using aggregate datasets may lead to different inferences and, hence, different policy implications (Carter et al. 2003). Provincial statistics of China have been extensively used, for example, in the influential papers of Lin (1992) on household responsibility system, and of Fan and Zhang (2002) on productivity and inequality. However, Herrmann-Pillath et al. (2002) suggested that provincial aggregates might not reflect the exact regional inequality of development for China and argued for uses of prefecture-level data. Chen and Huffman (2006) used a county-level dataset to investigate patterns of technical efficiencies in China's agriculture. Studies based on aggregate statistics could derive implications and recommendations about regional policies since levels of economic development and even the economic institutions vary across China (Krusekopf 2002). How such variation could affect agricultural efficiencies is a topic of interest and will yield critical implications for county-level agricultural policy.

In this paper we contribute to the existing literature by expanding our understanding of the factors that correlate with inefficiency in Chinese agriculture. Specifically, we examine the effects of a variety of production characteristics, investment variables, and government fiscal activity on agricultural production efficiency in China. The approach taken is a two-step procedure that combines non-parametric and parametric techniques. First, to estimate output-oriented measures of technical efficiency we use data envelope analysis (DEA), and secondly, a truncated regression model is used to identify inefficiency correlates where the significance of model estimates is inferred using a semi-parametric bootstrapping routine. Among the results, we find areas with a heavy industrial presence reduces agricultural production efficiency, and may be an indication that externalities from the industrial process, like air and ground water pollution, affects agricultural production. One somewhat surprising finding is that counties with a high ratio of credit outstanding relative to GDP are actually associated with greater inefficiency in agriculture. We also find evidence that counties where a large percentage of the rural labor force is engaged in agriculture-related work tend to be less efficient. This finding has the implication that policies to facilitate the removal of labor from agriculture but not necessarily from the rural labor force would bring about enhanced agricultural efficiency.

The rest of this paper is organized as follows. Section II formulates the econometric modeling strategy. Section III describes the data. Section IV presents the empirical results and includes a discussion of the implications of the findings. Section V outlines the method for controlling for influential observations and outlier and the corresponding results. Section VI concludes and summarizes.

II. Methodology:

When addressing issues related to technical efficiency, researchers generally adopt one of two commonly used methods. The first is the stochastic frontier approach which is amiable to a variety of statistical tools and estimation techniques. A limitation of the stochastic method is the requirement that the production technology be specified a priori. Imposing a specific functional form for a technology which is, in most cases, unknown can be problematic since alternative specifications can lead to different conclusions (Gong and Sickles 1992; Zhu, Ellinger and Shumway 1995; and Giannikas, Tran, and Tzouvelekas 2003). The second method involves non-parametric estimation of the production frontier which, rather than explicitly specifying a functional form, relies on more general assumptions typical of microeconomic production theory such as convexity, continuity, and free disposability to define the efficient frontier[1]. One criticism of non-parametric methods has been the lack of a clear link between non-parametric estimation of efficiency relationships and the subsequent statistical analysis used to explain variation thereon[2]. In fact, the lack of theoretical results to allow reasonable inference in two-stage procedures has resulted in the ad hoc use of various statistical models that does not necessarily follow from underlying process generating the data.[3] Fortunately the use of bootstrapping does provide an alternative means for statistical inference to proceed when more general results about the sampling properties of the estimators being considered are non-existent or intractable (Efron and Tibshirani, 1986). The algorithm described in Simar and Wilson (2007) details the data generating process (DGP) and subsequently the means by which inference might proceed when the objective is to describe variation in efficiency estimates obtained via non-parametric methods such as DEA. In this paper we use DEA to estimate output-oriented measures of technical efficiency in Chinese agriculture and identify the correlates of inefficiency in the second stage by using a truncated regression model with parameter significance inferred with the bootstrap algorithm detailed in Simar and Wilson (2007).

2.1 Data Envelope Analysis

For a representative firm operating at an inefficient point in the production set a measure of inefficiency is obtained by measuring the Euclidian distance from that point to the frontier. To measure technical inefficiency we compute output efficiency scores of the Farrell (1957) type, the reciprocal of the Shephard (1970) output distance function, and is obtained by solving the following for each firm i=1,…,n

$$
\begin{aligned}
\hat{\delta}_i = & \max_{\delta_i, \theta_1, \ldots, \theta_n} \delta_i \\
& st. \quad \delta_i \mathbf{y}_i \leq \mathbf{Y\theta} \\
& \quad \mathbf{x}_i \geq \mathbf{X\theta} \\
& \quad \sum \theta_i = 1 \\
& \quad \theta_i \geq 0 \quad \forall i = 1, \ldots, n
\end{aligned} \tag{1}
$$

where $\mathbf{y}_i$ and $\mathbf{x}_i$ are output and input vectors for the representative firm with k outputs and r inputs respectively, $\mathbf{Y} = [\mathbf{y}_1 \ldots \mathbf{y}_n]$ is a (k×n) matrix, $\mathbf{X} = [\mathbf{x}_1 \ldots \mathbf{x}_n]$ is a (r×n) matrix, and $\boldsymbol{\theta}$ is an (n×1) column of parameter values in the unit interval. For a particular firm, the parameter $\hat{\delta}_i$ is an estimate of the maximum amount by which output could be increased while using the same inputs[4].

2.2 Regression Analysis and Inference

Since the $\hat{\delta}_i$'s will necessarily be confined to the interval $\hat{\delta}_i \in [1, \infty)$ a truncated regression model is used to explain exogenous factors associated with inefficiency. Efficiency scores with a value of one indicate efficient operation (i.e. lies on the technology frontier) and values larger than one indicate a point that lies in the production set but is not on the frontier and represents technical inefficiency since the output could be increased by $\left[\hat{\delta}_i - 1\right] \times 100\%$ using the same amount of output. Of course, larger values represent an increasing degree of inefficiency. The relationship between the estimated efficiency scores and exogenous variables is represented via the following model

$$
\begin{aligned}
& \hat{\delta}_i = \mathbf{z}_i \boldsymbol{\beta} + \varepsilon_i \\
& \text{st. } \hat{\delta}_i \geq 1
\end{aligned} \tag{2}
$$

where vector $\mathbf{z}_i$ is a ($1 \times q$) collection of environmental variables exogenous to the firm, $\boldsymbol{\beta}$ is a ($q \times 1$) vector of parameters to be estimated, and ε_i is a continuous iid random error term. The $\mathbf{z}_i$'s are those variables that are used to explain the observed inefficiency. Given the requirement that δ_i be greater than or equal to one, the random error ε_i in (2) is assumed to be normally distributed with left-truncation at $\left(1-\mathbf{z}_i\boldsymbol{\beta}\right)$, and variance σ_ε^2. Given this specification, the **β**'s in equation (2) can be estimated by maximizing the following likelihood function

$$L=\prod_{i=1}^{n}\frac{1}{\sigma_\varepsilon}\phi\left(\frac{\hat{\delta}_i-\mathbf{z}_i\boldsymbol{\beta}}{\sigma_\varepsilon}\right)\left[1-\Phi\left(\frac{1-\mathbf{z}_i\boldsymbol{\beta}}{\sigma_\varepsilon}\right)\right]^{-1} \tag{3}$$

where $\phi(.)$ and $\Phi(.)$ are the probability density and cumulative distribution functions for the standard normal respectively. This is similar to the usual likelihood for a regression model except for the right-most bracketed term which ensures the density will integrate to unity after truncation.

In processes such as the one we described above, theoretical results regarding the precision of the parameters are either non-existent or, when such results do exist, only applicable to specific formulations (i.e. Gijbels et al. 1999). Fortunately so long as the DGP has been described, bootstrap techniques can be used to empirically approximate the sampling distributions of the parameters of interest and use these to infer the precision of these estimates (see Efron and Tibshirani 1986). The Simar and Wilson (2007) bootstrapping algorithm is summarized as follows[5]

[1] Estimate the efficiency scores using DEA to obtain $\hat{\delta}_i \; \forall i=1,...,n$.

[2] Obtain estimates $\hat{\boldsymbol{\beta}}$ and $\hat{\sigma}_\varepsilon$ by maximizing the log of the likelihood function in equation (3).

[3] Replicate steps [3.1]-[3.3] L times to produce a sequence of l=1,...,L bootstrap estimates $\left\{\boldsymbol{\beta}_l^*\right\}_{l=1}^{L}$. These estimates are used to approximate the sampling distribution for each parameter of interest.

[3.1] For each observation i=1,…,n, draw from the truncated normal described in (3) with standard deviation $\hat{\sigma}_\varepsilon$ to obtain a value d_i. To obtain this random draw, first compute a standardized value for each i, $sv_i = \frac{1-\mathbf{z}_i\hat{\boldsymbol{\beta}}}{\hat{\sigma}_\varepsilon}$, next draw v_i from the uniform distribution on the unit interval and use this to compute the draw $d_i = \Phi(sv_i) + [1-\Phi(sv_i)] * v_i$.

[3.2] Scale d_i to get the deviate from the left-truncated normal distribution, $u_i = \hat{\sigma}_\varepsilon * \Phi^{-1}(d_i)$, and use this to obtain the fitted value $\delta_i^* = \mathbf{z}_i\hat{\boldsymbol{\beta}} + u_i$.

[3.3] Perform operations [3.1]-[3.2] for each of the i=1,…,n firms to get a vector of estimates $\delta_i^* \;\forall i = 1,...,n$. Finally, using the fitted efficiency scores and the original explanatory variables in $\mathbf{z}_i$, use the method of maximum likelihood to maximize the truncated likelihood function in (3) to obtain a single set of bootstrap estimates $\boldsymbol{\beta}_l^*$.

The sequence of L bootstrap parameter estimates will empirically approximate the sampling distribution for each of the q parameters in $\boldsymbol{\beta}$. Creating a histogram using the sequence of bootstrap values for each parameter reveals an approximation of its distribution and can be used to determine whether or not a particular parameter was significantly different from zero at a given level of significance. For example, a (1-α)*100% confidence interval for a particular parameter β_q is found by ordering the L bootstrap estimates of β_q^* from lowest to highest and then remove the lowest (α/2)*L observations from both the lower and upper end of the sequence. Denote the lowest value in the remaining sample by $\underline{\beta}^*_{q,\frac{\alpha}{2}}$ and the largest remaining value by $\overline{\beta}^*_{q,\frac{\alpha}{2}}$, it follows that a (1-α)*100% confidence interval for β_q is given by $\left[\underline{\beta}^*_{q,\frac{\alpha}{2}}, \overline{\beta}^*_{q,\frac{\alpha}{2}}\right]$. For a particular level of significance α, if this interval does not include zero, then we reject the null hypothesis that the parameter β_q equals zero.

III. Description of Data and Efficiency Results

Agricultural efficiency scores are computed based on two outputs and four inputs. Agricultural output is measured by the amount of grain and meat produced and the four inputs are agricultural labor, mechanical power, fertilizer input, and sown area of agricultural crops. Data on agricultural output and inputs are based on a survey conducted by the Statistical Bureau of China for the year 1999. After removing observations with missing and unreasonable data, the final sample includes data from a total of 2,037 geographic areas or county equivalents. In 1999 the average county produced approximately 247,000 tons of grain and 22,000 tons of meat[6] (Table 1). The average number of agricultural workers was nearly 150 thousand and the average amount of mechanical power used in Watt equivalents was roughly 200 million, the equivalent of about 270 thousand horsepower hours. The average area of agricultural crop production was just over 70 thousand hectares to which an average of almost 19 thousand tons of fertilizer was applied.

Transforming the output and input variables using natural logs and estimating the output efficiency score for each of the counties using DEA resulted in an average efficiency score of 1.07 with 54 counties deemed to be operating efficiently with efficiency estimates equal to unity. Roughly speaking, these results imply that on average an additional 7% increase in output could be achieved with the same inputs. Note that since we included only variable inputs which are chosen by agricultural operators these output efficiency scores will not capture the differences in non-discretionary inputs such as land quality or climatic conditions that may also influence productivity. We partially control for characteristics such as these in the second stage regression where we use provincial indicator variables to capture regional variation in soil fertility and growing conditions.

It is common in the productivity literature to refer to factors that are beyond the control of the managers themselves as *environmental variables* and these might include, for example, location characteristics, labor market characteristics, ownership structure, and government regulations (Coelli, Rao, and Battese 1998; Fried, Schmidt, and Yaisawarng 1999). To explain the variation in the estimated output scores we propose the use of a number of different variables which we have grouped into three categories: i) production characteristics; ii) credit and government fiscal behavior; and iii) other local characteristics. The summary statistics for each of these categories is given in Table 2.

There are four variables included under the heading of production characteristics. The first three relate to the relative importance and structure of agriculture within each county and includes agriculture share of GDP, intensity of mechanical power usage in agriculture, and livestock share of total agricultural output. Agricultural production makes up about 35% of total output for the average county and ranges from 0.3% to just over 90%. The variable for intensity of mechanical relative to labor input in agriculture, computed as the ratio of mechanical power relative to agricultural labor, is an indicator of the relative importance of mechanization in production. This variable reveals significant variation in the intensity with which these types of inputs are used and ranges from counties that are relatively dominated by labor (minimum of 0.03) to counties that are relatively highly mechanized (maximum 280). An interesting policy issue to address is how areas specializing in livestock production compare to those that are dominated by crop production in terms of efficiency. To evaluate this we consider livestock share of output relative to total agricultural output. On a per ton basis, the average county has 10% of its agricultural output coming from livestock production and ranges from less than 1% to 96%. The fourth measure is agricultural GDP per agricultural worker and is intended to capture factors such as human capital and skill of agricultural labor.

Research has shown that credit considerations can influence agricultural efficiency in developing countries (Dong and Putterman 1997; Abdulai and Huffman 2000; Chavas et al. 2005). The ratio of credit outstanding to GDP has an average of 0.73 and is expected to give an indication as to the indebtedness of the county in general and how this relates to efficient production in the agricultural sector. Ranging from less than 0.05 to over 7.7, some counties have very little credit outstanding while in others it is several times that of income. Two variables were created to assess whether government fiscal activity impacts efficiency. First, fiscal revenue relative to expenditures as a measure of the fiscal balance is used to inform whether inefficiency is influenced by government deficit spending. Fiscal revenue relative to expenditure averaged about 0.56 indicating that the average local county government body spends roughly twice what they collect. Secondly, we include fiscal expenditures relative to GDP to measure how the relative size of government spending relative to total income within the county affects efficiency. Fiscal expenditures relative to GDP captures government activities (which would include investment, services, and government employment) relative to total production in the region. Here, average spending was slightly more than ten percent of GDP.

County characteristics include variables related to industrial structure, income levels, and demographic composition. Industry share of total GDP was found to be approximately 35% (about the same as agricultural share of GDP) and ranges from a low of approximately 4% to nearly 86%. We also include a measure of income for the non-agricultural population, non-agricultural GDP per non-agricultural worker, as a measure of affluence, skill, and level of development in the county outside agriculture. The data reveal that the representative county population is 82% rural. Research by Yao and Liu (1998) found that a higher share of population that is rural and a larger share of the agricultural workforce devoted to grain production each contributed to inefficiency in Chinese agriculture. To get a better understanding of how "ruralness" might affect efficiency of agricultural production we include the rural share of the population. We also include agriculture's share of the rural labor force to inform on how rural employment dominated by agriculture might influence agricultural efficiency locally, if indeed at all. The data reveal that agriculture is the main employer of rural labor with an average share of 0.75 and ranging from about 13% to 100%.

IV. Results

To explain variation in non-parametric output efficiency scores estimated using DEA, we specify a truncated regression model with left truncation at unity. While the truncated regression model is arguably a more appropriate model than limited dependent variable models such as the Tobit[7], truncation at one necessarily means that those observations for counties which had an estimated efficiency score equal to one, i.e. those that are efficient, will be lost. However since only 54 counties having an output efficiency score equal to one, this is not a major limitation since we are still left with more than 97% of our original sample. Given the explanatory variables mentioned in the previous section it is likely that some of these variables are highly collinear and this should be taken into consideration when conducting regression analysis to avoid problems associated with multicolinearity. The correlations matrix in Table 3 was used to guide the choice of explanatory variables used in the regression analysis to avoid including variables that are highly correlated.

The results of the regression analysis are presented in Table 4 and Table 5 where the latter includes the same regressors as the former except that unobservable provincial effects are captured through the inclusion of 29[8] provincial dummy variables (the default is Sichuan province with 144 usable observations). Notice that since the dependent variable, output efficiency score, is a unit free measurement there is no a priori reason to expect there exists any mechanical correlation with the explanatory variables in the truncated regression. That is to say, there is no reason to assume that endogeneity is a problem.

In columns 1 and 4 of Table 4 the parameter estimate for agricultural share of GDP is negative and statistically significant (at the 1% level). Notice that, given our DEA formulation in equation (1), increasing variables with a *positive* estimated regression coefficient results in greater *inefficiency,* whereas an increase in a variable with a *negative* coefficient would correspond to greater technical *efficiency*. Thus, the negative sign for the coefficient estimates corresponding to agricultural share of GDP imply that areas that are more specialized in agricultural production tend to be more technically efficient. Greater efficiency where agriculture is relatively more important, at least in terms of total county product, might partially be due to greater competition among producers and a higher likelihood that producers will be able to copy or imitate the best production practices of others in the area. Additionally, a county where agriculture is important may also expend more resources catering to this industry in terms of service provision, diffusion of best practice technologies, and engaging in activities that otherwise promote efficient production of agricultural goods. There is some evidence to suggest that counties that are mechanized are less efficient (column 1, Table 4). However, when provincial controls are added there is evidence to the contrary (column 4, Table 5). Considering the role of livestock intensity in agricultural output, the third production characteristic, the results indicate that higher share of total agricultural output from livestock is associated with greater *inefficiency* and statistically significant in all specifications, both without and with provincial controls. One possible explanation might be related to the fact that energy is lost when livestock convert feed into tissue and low conversion ratios might not be fully reflected in market prices. The final agricultural production characteristic, agricultural GDP per agriculture worker, appears to correlate *positively* with efficiency given the negative coefficient both without and with provincial controls as shown in column 1 of tables 4 and 5 respectively.

The inclusion of the credit variable was thought to provide insight into how the local investment environment might influence efficiency in agriculture. The estimated coefficient for credit to GDP ratio is positive and significant (at the 1% level) in all specifications that include this variable (columns 2, 4, 5, 6, and 7), regardless of the inclusion/exclusion of provincial controls (Table 4 and 5), indicating that counties where outstanding credit is large relative to total income tend to be associated with greater agricultural production inefficiency. While difficult to identify precisely the underlying causes, it may be an indication of excessive investment and that reducing the amount of investment capital by means of increasing consumption might increase efficiency. In addition, if the investment happened to be in primarily agriculture-related technologies unfamiliarity with these new technologies might be the cause of inefficiency for a short time just after adoption while the operators "learn" how to best use the given technology.

The ratio of fiscal revenue relative to expenditures, one of the two fiscal variables, has an estimated coefficient that is negative and is significant at the 5% level with provincial controls (column 2, Table

5). Recall that the average ratio of revenues relative to expenditures was approximately 0.55 (Table 2). If the local governments were to act in a more fiscally responsible manner and increase revenues to more closely match expenditures, these results suggest that improvements in agricultural efficiency might be possible. However, when provincial controls are not included, the estimated coefficient for fiscal revenue relative to expenditures is not longer significant (column 2, Table 5). At the same time, we also find higher government expenditure relative to GDP, the results of which are also robust to alternative specifications (columns 2, 4, 5, and 7) and the exclusion/inclusion of provincial controls, is correlated with increased *inefficiency* which would suggest reducing the relative size of government activity would improve efficiency. It has been shown that market distortions resulting from institutional framework contribute to inefficiency (Wang, Cramer, and Wailes 1996). To the extent that government fiscal activities distort market signals is one possible explanation for these results.

Industrial production share of GDP considers how industrial presence impacts the efficiency of agricultural production. A priori, a strong industrial sector could potentially improve or impede agricultural efficiency. For example, a large industrial presence might help by providing beneficial technology spillovers, better access to utilities, and improved transportation infrastructure. In contrast, industrial activity might hinder agricultural production due to negative externalities arising from the production process as well as this industry competing with the agricultural sector for land, labor, and investment capital. The results indicate a positive and statistically significant (at the 1% level) relationship between share of GDP from industry and output efficiency scores for the specifications in columns 3, 5, and 7 in both tables 4 and 5. Taken together, these results indicate that counties with a large share of total county product from industrial production have lower agricultural efficiency and suggest that any positive spillover effects between industrial and agricultural production, if they exist, are being overwhelmed by the negative ones. In recent years pollution resulting from industrial activity has become a concern in China and the current finding may be an indication of the adverse impact this is having on agricultural production. Another possibility is that these findings might also reflect a lower quality or intensity of agricultural labor in areas where there is heavy industrial production. Since we do not have any information that allows for the quality or the skill of agricultural workers, it is possible that competition from industrial producers has drawn skilled workers away from agriculture. An alternative explanation might point to local growth and development policies that favor industrial development but at the same time actually makes agriculture production more difficult.

The estimated coefficient for non-agricultural GDP per non-agricultural adult worker is negative and significant (at the 1% level) for all specifications that include this variable (columns 3 and 6 in Table 4 and 5). Capturing in part the role of human capital and skill of the non-agricultural workforce, GDP per non-agricultural working adult is negatively correlated with inefficiency and highly statistically significant. The implications of the relative size of the local population that is rural are mixed. In column 6 of Table 4 the estimated coefficient is positive and significant at the 1% level and suggests the larger is the percentage of the population that is rural, the greater is the inefficiency, a finding in agreement with work by Yao and Liu (1998). However, when provincial controls are used, there is evidence that the larger is the rural share of the population the more efficient is agricultural production (column 3, Table 5), a result significant at the 10% level. When considering how efficiency is affected by the composition of rural employment, we find counties with a *higher* share of agricultural labor relative rural labor tend to be *less efficient* for all specifications that include this variable (columns 3, 5, and 7, Table 4 and 5) and is significant at the 1% level in all specifications except one (column 5, Table 4) where it is significant at the 5% level. These findings may be an indication that an overly large share of the rural labor force is engaged in agricultural production and that policies to improve efficiency in the agricultural sector might involve drawing agricultural labor in rural areas into non-agricultural sectors in rural areas.

V. Controlling for Influential Observations and Outliers

Two criticisms often offered against the use of non-parametric methods to assess efficiency-related problems are: lack of statistical tools to evaluate the precision of parameter estimates in two-stage methods, an issue addressed in the previous sections; and the sensitivity of non-parametric methods to outliers. Not limited to non-parametric methods alone, the second has implication for both stochastic and non-parametric procedures but tend to be more problematic for non-parametric analyses. In non-parametric procedures like DEA, efficiency scores could be biased (upwards given our specification) if outliers are used to form the convex hull of the production set since the efficiency scores of the remaining observations are computed based on their relative location the frontier.

Visual methods can be useful when identifying outliers and influential observations when the sample size is small but when the sample size is rather large, such as with the dataset used in this paper, other methods must be used. As a robustness check to see if the results discussed in the previous section are sensitive to influential observations and outliers we compute leverage score (l_i) for each observation[9]. Computing the leverage score for each observation involves first estimating an efficiency score for each element of the dataset. Next, in jackknife-like approach we systematically exclude one observation from the dataset and re-compute the efficiency scores for the remaining observations (δ_j^l $\forall j = 1,..,i-1,i+1,...n$) and compare these new scores against the originals. Specifically, for each observation the leverage score is computed as follows

$$l_i = \left(\frac{\sum_{j=1, j\neq i}^{n} \left(\delta_j^l - \hat{\delta}_j \right)^2}{n-1} \right)^{\frac{1}{2}} \tag{4}$$

This procedure is repeated for the n-1 remaining observations to determine which observations exert influence or might be potential outliers. When a particular observation is used to define the convex hull of the production frontier and it has few or no close peers, then that observation might be said to be an influential one since the scores of other, non-efficient observations in the neighborhood are measured relative to that particular point. Further, if the output or input data for a particular observation happens to be measured with error so that it appears more “efficient” than it aught (i.e. reporting fewer inputs or inflated output), any observation whose performance is measured relative to that particular outlier will appear more inefficient than they actually are. By computing a leverage score as per equation (4) it is possible to identify observations that exert significant influence over the efficiency scores of other observations. A larger leverage score indicates a more influential, and possibly problematic, element of the dataset. In addition to outliers that influence the efficiency scores of its peers, an observation that is too far to the interior of the production set might be measured with error (i.e. output(s) being under reported or input usage over reported). These outliers to the interior of the production set can be identified by a leverage score that is very low or zero. Identifying influential observations and outliers and removing these from the dataset, the two-step estimation and inference procedure can be repeated to determine if results are robust to the removal of influential and outlier observations.

The methodology followed in this paper to test for the robustness of results to outliers and influential observations involves computing the leverage score for each element in the full dataset and then repeating the two-stage procedure outlined earlier having removed those elements with unusually high and low leverage scores. The rule used here is to remove five percent of each of the highest and lowest leverage scoring observations which leaves a sample size that is nine-tenths of the original. The production and input summary statistics for remaining observations and the resulting efficiency scores are presented in Table 6 and the summary statistics for the explanatory variables used in the second stage in Table 7.

The results in Table 8 correspond to those in Table 5 when no provincial controls are used and Table 9 is the counterpart to Table 6 when provincial dummy variables are included. Parameter significance is again determined based on the bootstrapping routine outlined earlier. A comparison of these results with those presented in the earlier section leads us to conclude that outliers are unlikely to be a factor here given the robustness of the results to the exclusion of outliers.

VI. Conclusions

To facilitate continued growth in a sustainable manner, policy to promote greater efficiency of agricultural production will become increasingly important in a bid to limit the strain on the country's environment, water resources, and infrastructure. Given roughly 200 million small-scale households in China, policy promoting increased efficiency in agricultural production would come at a key point in time when the country is undergoing a significant change in the composition of its rural-urban population as the urban population in China expected to grow significantly during the next few decades. During this period of transition policy makers have a unique opportunity to develop a strategy and implement changes that will facilitate greater productive efficiency, thereby limiting at least some of the difficulties that are sure to accompany transition towards urbanization and the increasing dependence on the rural hinterland to supply agricultural goods that follows.

This study of technical efficiency in Chinese agricultural production involved a two-stage process where output efficiency scores were estimated using DEA and variation in the resulting efficiency scores was explained by using a truncated regression model. Where such two-stage methodologies have been criticized for lacking the proper theoretical results to conduct statistical inference in the past, we circumvent this problem by relying on a semi-parametric bootstrapping routine to conduct inference. Using a dataset consisting of a cross-section of more than 2,000 counties revealed a number of characteristics that correlate with agricultural inefficiency.

Among the results, we find that counties that are heavily vested in agriculture tend to be more efficient while those with a high industrial presence tend to be less efficient. This latter result leads us to suppose that complementarities are limited between agricultural and industrial production processes and that there exist potentially negative spillovers from industrial production that hinders agricultural production such as pollution. In recent years, the environmental consequence of pollution arising from rapid and unchecked industrialization has become an important topic of debate among researchers, regulators, and environmentalists. Given the increasing role environmental considerations are likely to take in the future, the relationship between industrial activity, pollution, and agricultural production is one topic that deserves further consideration.

Additional findings with policy implications include those related to the credit and government fiscal variables. One somewhat surprising finding is that counties with a high ratio of credit outstanding relative to GDP might be actually associated with greater inefficiency in agriculture. While there are a number of possible explanations, one might be an indication of excessive investment in which case reducing the amount of investment capital by means of increasing consumption might increase efficiency. It is well known that the savings rate in China is considerably higher than most other countries, and certainly higher than those in the West. Indeed, limiting locally available credit through reduced savings would encourage greater consumption and reduce the rate of capital accumulation which could slow the pace of growth. A slowed pace of growth might allow for new technologies to be adopted at a slower rate which would allow for potentially greater efficiency of use by allowing its users to become more familiar with these technologies before ushering in new ones. On the part of local governments, we find evidence efficiency could be improved by reducing expenditures as a fraction of aggregate county income as well as relative to revenues collected. When devising optimal policy it is important to note that while such actions may result in greater efficiency of agricultural production though the reduction of superfluous program spending, limiting government expenditures may also reduce the level of service provision and public investment which might limit productive capacity in the future.

Finally, we also find evidence that the larger is agriculture's share of rural labor the less technically efficient is agricultural production in the county. This finding has the implication that policy to facilitate the removal of labor from agriculture, but not necessarily from the rural labor force, would bring about enhanced agricultural efficiency. An appropriate set of policies here would involve promoting non-agricultural rural development as opposed to facilitating migration to cities. While it is beyond the scope of this paper to make definitive recommendations on migration policy, such a finding does call into question policies that promote wholesale migration of rural residents to urban centers when economic improvement might also be achieved by developing the appropriate rural sectors.

[1] See Tulkens and Eeckaut (1995) for further discussion of these assumptions and their implications for analysis using non-parametric techniques.

[2] An exception is Gijbels et al. (1999) who provide asymptotic sampling results for the specific one-input and one-output case.

[3] See Simar and Wilson (2007) for a review of this literature.

[4] The optimization problem specified here allows for constant returns to scale in the lower range of inputs up to a point and then decreasing returns (Tulkens and Eeckaut 1995)

[5] Other studies that have used this bootstrapping procedure in empirical applications include Oliveira and Santos (2005) and Kuosmanen, Pemsl, and Wesseler (2006).

[6] Beef, pork, and mutton.

[7] See Simar and Wilson (2007) for a discussion of the appropriateness of various regression models used to explain variation in efficiency scores estimated using DEA.

[8] Tibet and Taiwan are excluded.

[9] The method used here follows Sampaio De Sousa and Stosic (2005, p.162)

VII. References:

Abdulai, A. and W. Huffman, "Structural Adjustment and Economic Efficiency of Rice Farmers in Northern Ghana." *Economic Development and Cultural Change* (2000):503-520.

Carter C.J., Chen J., & Chu B., "Agricultural productivity growth in China: Farm level versus aggregate measurement," *China Economic Review* 14 (2003): 53-71.

Chavas, J-P, R. Petrie R., and M. Roth, "Farm Household Production Efficiency: Evidence from the Gambia," *American Journal of Agricultural Economics* 87(2005):160-179.

Chen, Z. and W. Huffman, "Measuring County-Level Technical Efficiency of Chinese Agriculture: a Spatial Analysis," Book Chapter, included in *China's Agricultural Development*, Edited by X-Y. Dong, S. Song, & X. Zhang, Ashgate Publishing Limited, UK (2006) pp.152-172.

Chen, Z, W. Huffman, and S. Rozelle, "Farm Technology and Technical Efficiency: Evidence from Four Regions in China," Iowa State University, Department of Economics Working Paper # 06022. (May 2006) http://www.econ.iastate.edu/research/webpapers/paper_12605_06022.pdf

Coelli, T., Rao, D.S.P., Battese, G.E., An Introduction to Efficiency and Productivity Analysis. Kluwer Academic Publishers, Inc., Boston (1998).

Dong, X. and L. Putterman, "Productivity and Organization in China's Rural Industries: A stochastic Frontier Analysis," *Journal of Comparative Economics* 24 (1997):181-201.

Efron, B. and R. Tibshirani, "Bootstrap Methods for Standard Errors, Confidence Intervals, and Other Measures of Statistical Accuracy," *Statistical Science* 1 (1986):54-75.

Fan, S. and X. Zhang, "Production and Productivity Growth in Chinese Agriculture: New National and Regional Measures," *Economic Development and Cultural Change* (2002):819-838.

Farrell, M., "The Measurement of Productive Efficiency," *Journal of the Royal Statistical Service*, Series A: General 120 (1957):253-281.

Fried, H.O., S.S. Schmidt, and S. Yaisawarng, "Incorporating the Operating Environment Into a Nonparametric Measure of Technical Efficiency," *Journal of Productivity Analysis* 12 (1999):249-267.

Gale, F., "China's Food and Agriculture: Issues for the 21st Century," Market and Trade Economics Division, Economic Research Service, U.S. Department of Agriculture. Agriculture Information

Bulletin No. 775 (2002).

Giannikas, D., K. Tran, and V. Tzouvelekas, "On the Choice of Functional Form in Stochastic Frontier Modeling," *Empirical Economics* 28 (2003):75-100.

Gijbels, B., E. Mammen, B. Park, and L. Simar, "On Estimation of Monotone and Concave Frontier Functions," *Journal of the American Statistics Association* 94 (1999): 220-228.

Gong, B., and R. Sickles, "Finite Sample Evidence on the Performance of Stochastic Frontiers and Data Envelopment Analysis Using Panel Data," *Journal of Econometrics* 51 (1992):259-284.

Herrmann-Pillath C., Kirchert D., & Pan J. Prefecture-level statistics as a source of data for research into China's regional development. *The China Quarterly* 172 (2002): 956-985

Krusekopf C.C., "Diversity in Land-Tenure Arrangements under the Household Responsibility System in China," *China Economic Review* 13 (2002): 297-312.

Kumbhakar S.C., and C. Lovell, *Stochastic Frontier Analysis.* London: Cambridge University Press (2000).

Kuosmanen, T., D. Pemsl, and J. Wesseler, "Specification and Estimation of Production Functions Involving Damage Control Inputs: A Two-Stage, Semiparametric Approach," *American Journal of Agricultural Economics* 88 (2006):499-511.

Lin J.Y., "Rural Reforms and Agricultural Productivity Growth in China," *American Economic Review* 82 (1992):34-51.

Liou M. and L.Yu, "Assessing Statistical Accuracy in Ability Estimation: A Bootstrap Approach," *Psychometrika* 56 (1991):55-67.

Lu Z. and S. Song, "China's Regional Disparities in 1978-2000," *Current Politics and Economics of Asia* 13 (2004):253-279.

Mao W.N. and W. Koo, "Productivity Growth, Technological Progress, and Efficiency Change in Chinese Agriculture after Rural Economic Reforms: A DEA Approach," *China Economic Review* 8 (1997):157-74.

Meeusen W. and V. Broeck, "Efficiency Estimation from Cobb-Douglas Production Functions with Composed Error," *International Economic Review* 18 (1977):435-44.

Oliveira, M. and C. Santos, "Assessing School Efficiency in Portugal Using FDH and Bootstrapping," *Applied Economics* 37 (2005):957-968.

Sampaio De Sousa, M., and B. Stosic, "Technical Efficiency of the Brazilian Municipalities: Correcting Nonparametric Frontier Measurements for Outliers," *Journal of Productivity Analysis* 24(2005):157-181.

Shephard, R.W., *Theory of Cost and Production Function*. Princeton University Press, Princeton (1970).

Simar, L. and P. Wilson, "Estimation and Inference in Two-Stage, Semi-Parametric Models of Production Processes," *Journal of Econometrics* 136 (2007):31-64.

State Statistical Bureau. County-level demographic and economic data: 1992, 1995, 1999. http://www.stats.gov.cn/tjsj/qtsj/index.htm (accessed March 2006).

Tian, W. and G. Wan, "Technical Efficiency and Its Determinants in China's Grain Production," *Journal of Productivity Analysis* 13 (2000):159-174.

Tulkens, H. and P. Eeckaut, "Non-Parametric Efficiency, Progress and Regress Measures for Panel Data: Methodological Aspects," *European Journal of Operational Research* 80 (1995):474-499.

Wilson, P., "Testing Independence in Models of Productive Efficiency," *Journal of Productivity Analysis* 20 (2003):361-390.

Wan G.H., and E. Cheng, "Effects of Land Fragmentation and Returns to Scale in the Chinese Farming Sector," *Applied Economics* 33 (2001):183-94.

Wang, J., E. Wailes, E.J., and G. Cramer, "A Shadow-Price Frontier Measurement of Profit Efficiency in Chinese Agriculture," *American Journal of Agricultural Economics* 78 (1996):146-156.

Wang J.R., G. Cramer, and E. Wailes, "Production Efficiency of Chinese Agriculture: Evidence from Rural Household Survey Data," *Agricultural Economics* 15 (1996):17-28.

Yang H., "China's Maize Production and Supply from a Provincial Perspective," Working paper No. 96/7 Center for Asian studies, Chinese economy research unit, The University of Adelaide (1996).

Yao, S. and Z. Liu, "Determinants of Grain Production and Technical Efficiency in China," *Journal of Agricultural Economics* 49 (1998):171-184.

Yao S, Z. Liu, and Z. Zhang, "Spatial Difference of Grain Production Efficiency in China, 1987-1992," *Economics of Planning* 34 (2001):139-157

Zhang X., S. Fan, L. Zhang, and J. Huang, "Local Governance and Public Goods Provision in Rural China," *Journal of Public Economics* 88 (2004):2857-2871.

Zhu, S., P. Ellinger, and C.R. Shumway, "The Choice of Functional Form and Estimation of Banking Efficiency," *Applied Economic Letters* 2 (1995):375-279.

Table 1. First Stage Summary Statistics – Complete Sample

		Obs	Mean	Std. Dev.	Min	Max
Output	Grain (ton)	2037	246669	234267	132	2248891
	Meat (Beef, Pork, Mutton - ton)	2037	21920	20977	51	189090
Input	Agricultural labor (1000)	2037	149	114	1	688
	Mechanical Power (1,000, 000 Watt)	2037	212	221	1	2007
	Fertilizer (ton, weighted)	2037	18661	18495	4	150493
	Sown area of agricultural crops (Hectare)	2037	70640	51010	103	320021
	DEA Output Score (No. efficient = 54)	2037	1.070	0.044	1	1.578

Table 2. Explanatory Variables Summary Statistics – Complete Sample

		Obs.	Mean	Std. Dev.	Min	Max
Production Characteristics						
	Agr. share of GDP	1983	0.350	0.151	0.003	0.908
	Agr. mechanical power intensity	1983	1.891	6.516	0.098	280.000
	Livestock share of agr. output	1983	0.119	0.153	0.007	0.960
	Agr. GDP per agr. worker	1983	5179.570	4231.238	12.075	84535.000
Credit and Fiscal						
	Credit relative to GDP	1983	0.733	0.461	0.045	7.705
	Fiscal revenue relative to expenditures	1983	0.556	0.220	0.025	1.861
	Fiscal expenditures relative to GDP	1983	0.102	0.081	0.019	1.005
Other Local						
	Industrial GDP relative to GDP	1983	0.347	0.139	0.039	0.859
	Non-agr. GDP per adult non-agr. worker	1983	6597.053	5868.538	451.037	98612.720
	Rural share of population	1983	0.823	0.135	0.075	1.000
	Agr. labor share of rural labor	1983	0.749	0.135	0.226	1.000

Table 3. Explanatory Variables Correlation

	AGDP/ TGDP	Mach/ ALab	Meat/ Grain	ln(AGDP/ ALab)	Credit/ TGDP	FRev/ FExp	FRev/ TGDP	IGDP/ TGDP	ln(NAGDP/ NAPop)	RPop/ TPop	ALab/ RLab
AGDP/TGDP	1.0000										
Mach/ALab	-0.0528	1.0000									
Meat/Grain	0.0390	-0.0001	1.0000								
ln(AGDP/ALab)	-0.0155	0.2390	0.0381	1.0000							
Credit/TGDP	0.0102	0.0139	-0.0228	-0.1053	1.0000						
FRev/FExp	-0.4406	0.0359	-0.0762	0.3302	-0.1745	1.0000					
FRev/TGDP	0.3131	-0.0051	0.1120	-0.3822	0.2223	-0.6516	1.0000				
IGDP/TGDP	-0.8459	0.0397	-0.0811	0.0669	-0.0946	0.4894	-0.4267	1.0000			
ln(NAGDP/NAPop)	-0.7565	0.1017	-0.0218	0.4532	-0.2252	0.5908	-0.5488	0.7239	1.0000		
RPop/TPop	0.2117	-0.2241	-0.0622	-0.4893	-0.2938	-0.1002	-0.0135	-0.0919	-0.1914	1.0000	
ALab/RLab	0.5265	-0.0191	0.0700	-0.2654	0.1311	-0.4210	0.4260	-0.5283	-0.4786	-0.1126	1.0000

Table 4. Truncated Regression Results – Complete Sample

Variable		#1 [a]	#2	#3	#4	#5	#6	#7
Production Characteristics								
	Agr. share of GDP	-0.051 ***			-0.081 ***			
	Agr. mechanical power intensity (x100)	0.028 *			-0.033			-0.026
	Livestock share of agr. output	0.071 ***			0.072 ***			0.079 ***
	Agr. GDP per agr. worker	-0.034 ***						
Credit and Fiscal								
	Credit relative to GDP		0.014 ***		0.012 ***	0.013 ***	0.021 ***	0.014 ***
	Fiscal revenue relative to expenditures		0.009					
	Fiscal expenditures relative to GDP		0.147 ***		0.155 ***	0.177 ***		0.148 ***
Other Local								
	Industrial GDP relative to GDP			0.165 ***		0.087 ***		0.080 ***
	Non-agr. GDP per adult non-agr. worker			-0.030 ***			-0.008 ***	
	Rural share of population			0.016			0.041 ***	
	Agr. labor share of rural labor			0.045 ***		0.024 **		0.021 ***
	Constant	1.355 ***	1.031 ***	1.211 ***	1.060 ***	0.986 ***	1.076 ***	0.986 ***
	Regional Controls [b]	no	no	no	no	no	no	no
	n	1983	1983	1983	1983	1983	1983	1983
	log likelihood	3872.8	3649.5	3674.84	3757.5	3679.5	3633.6	3740.4
	Wald Chi sq	588.33	125.48	163.66	358.39	184.71	90.24	322.47

[a] Inference is based on confidence intervals obtained from 1000 bootstrap iterations. The superscripts: *, **, and *** indicates the value zero does not fall within the 90, 95, and 99 percent confidence intervals respectively.
[b] Regional controls consist of 29 provincial dummy variables with Sichuan province being the default province.

Table 5. Truncated Regression Results – Complete Sample

Variable		#1 [a]	#2	#3	#4	#5	#6	#7
Production Characteristics								
	Agr. share of GDP	-0.032 ***			-0.050 ***			
	Agr. mechanical power intensity (x100)	0.009			-0.033 *			-0.026
	Livestock share of agr. output	0.066 ***			0.066 ***			0.071 ***
	Agr. GDP per agr. worker	-0.031 ***						
Credit and Fiscal								
	Credit relative to GDP		0.011 ***		0.010 ***	0.011 ***	0.012 ***	0.011 ***
	Fiscal revenue relative to expenditures		-0.017 **					
	Fiscal expenditures relative to GDP		0.075 ***		0.107 ***	0.113 ***		0.092 ***
Other Local								
	Industrial GDP relative to GDP			0.067 ***		0.038 ***		0.038 ***
	Non-agr. GDP per adult non-agr. worker			-0.015 ***			-0.006 ***	
	Rural share of population			-0.019 *			-0.004	
	Agr. labor share of rural labor			0.046 ***		0.037 ***		0.039 ***
	Constant	1.285 ***	1.021 ***	1.115 ***	1.021 ***	0.967 ***	1.073 ***	0.960 ***
	Regional Controls [b]	yes	yes	yes	yes	yes	yes	yes
	n	1983	1983	1983	1983	1983	1983	1983
	log likelihood	4133.2	3952.9	3937.25	4037.5	3959.9	3931.6	4028.3
	Wald Chi sq	1291	741.69	697.79	1001.6	759.03	689.86	974.07

[a] Inference is based on confidence intervals obtained from 1000 bootstrap iterations. The superscripts: *, **, and *** indicates the value zero does not fall within the 90, 95, and 99 percent confidence intervals respectively.
[b] Regional controls consist of 29 provincial dummy variables with Sichuan province being the default province.

Table 6. First Stage Summary Statistics – High/Low Leverage Observations Removed

		Obs	Mean	Std. Dev.	Min	Max
Output	Grain (ton)	1832	247602	236301	135	2099654
	Meat (Beef, Pork, Mutton - ton)	1832	21880	20983	51	189090
Input	Agricultural labor (1000)	1832	149	116	1	688
	Mechanical Power (1,000, 000 Watt)	1832	213	220	1	2007
	Fertilizer (ton, weighted)	1832	18789	18819	4	150493
	Sown area of agricultural crops (Hectare)	1832	70680	51746	187	320021
	DEA Output Score (No. efficient = 53)	1832	1.069	0.044	1	1.568

Table 7. Explanatory Variables Summary Statistic – High/Low Leverage Observations Removed

Production Characteristics		Obs	Mean	Std. Dev.	Min	Max
	Agr. share of GDP	1779	0.351	0.152	0.003	0.908
	Agr. mechanical power intensity	1779	1.899	6.818	0.098	280.000
	Livestock share of agr. output	1779	0.118	0.130	0.007	2.603
	Agr. GDP per agr. worker	1779	5146.726	3872.154	12.075	34784.000
Credit and Fiscal						
	Credit relative to GDP	1779	0.744	0.475	0.080	7.705
	Fiscal revenue relative to expenditures	1779	0.554	0.221	0.025	1.861
	Fiscal expenditures relative to GDP	1779	0.103	0.083	0.019	1.005
Other Local						
	Industrial GDP relative to GDP	1779	0.345	0.139	0.039	0.859
	Non-agr. GDP per adult non-agr. worker	1779	6525.386	5828.693	451.037	98612.720
	Rural share of population	1779	0.821	0.138	0.075	1.000
	Agr. labor share of rural labor	1779	0.750	0.135	0.226	1.000

Table 8. Truncated Regression Results – High/Low Leverage Observations Removed

Variable		#1 [a]	#2	#3	#4	#5	#6	#7
Production Characteristics								
	Agr. share of GDP	-0.054 ***			-0.083 ***			
	Agr. mechanical power intensity (x100)	0.025			-0.035			-0.003
	Livestock share of agr. output	0.087 ***			0.096 ***			0.106 ***
	Agr. GDP per agr. worker	-0.035 ***						
Credit and Fiscal								
	Credit relative to GDP		0.015 ***		0.013 ***	0.014 ***	0.021 ***	0.014 ***
	Fiscal revenue relative to expenditures		0.011					
	Fiscal expenditures relative to GDP		0.159 ***		0.151 ***	0.186 ***		0.146 ***
Other Local								
	Industrial GDP relative to GDP			0.189 ***		0.095 ***		0.083 ***
	Non-agr. GDP per adult non-agr. worker			-0.034 ***			-0.009 ***	
	Rural share of population			0.019 *			0.046 ***	
	Agr. labor share of rural labor			0.051 ***		0.027 **		0.022 ***
	Constant	1.367 ***	1.025 ***	1.232 ***	1.055 ***	0.976 ***	1.080 ***	0.978 ***
	Regional Controls [b]	no	no	no	no	no	no	no
	n	1779	1779	1779	1779	1779	1779	1779
	log likelihood	3475.5	3255.9	3289.69	3362.2	3284.1	3243.4	3348.2
	Wald Chi sq	572.67	115.6	167.88	344.93	170.34	87.14	315.37

[a] Inference is based on confidence intervals obtained from 1000 bootstrap iterations. The superscripts: *, **, and *** indicates the value zero does not fall within the 90, 95, and 99 percent confidence intervals respectively.
[b] Regional controls consist of 29 provincial dummy variables with Sichuan province being the default province.

Table 9. Truncated Regression Results – High/Low Leverage Observations Removed

Variable		#1 [a]	#2	#3	#4	#5	#6	#7
Production Characteristics								
	Agr. share of GDP	-0.033 ***			-0.049 ***			
	Agr. mechanical power intensity (x100)	0.004			-0.034 *			-0.029
	Livestock share of agr. output	0.082 ***			0.090 ***			0.010 ***
	Agr. GDP per agr. worker	-0.033 ***						
Credit and Fiscal								
	Credit relative to GDP		0.012 ***		0.011 ***	0.012 ***	0.013 ***	0.012 ***
	Fiscal revenue relative to expenditures		-0.016 **					
	Fiscal expenditures relative to GDP		0.080 ***		0.103 ***	0.119 ***		0.089 ***
Other Local								
	Industrial GDP relative to GDP			0.081 ***		0.044 ***		0.042 ***
	Non-agr. GDP per adult non-agr. worker			-0.018 ***			-0.006 ***	
	Rural share of population			-0.017			0.000	
	Agr. labor share of rural labor			0.051 ***		0.040 ***		0.041 ***
	Constant	1.296 ***	1.017 ***	1.120 ***	1.015 ***	0.959 ***	1.069 ***	0.949 ***
	Regional Controls [b]	yes	yes	yes	yes	yes	yes	yes
	n	1779	1779	1779	1779	1779	1779	1779
	log likelihood	3720.9	3536.3	3524.27	3625.9	3544.1	3517	3619.5
	Wald Chi sq	1263	679.72	643.79	962.41	698.66	632.38	941.08

[a] Inference is based on confidence intervals obtained from 1000 bootstrap iterations. The superscripts: *, **, and *** indicates the value zero does not fall within the 90, 95, and 99 percent confidence intervals respectively.
[b] Regional controls consist of 29 provincial dummy variables with Sichuan province being the default province.