Appendix: Regression Analysis

China

Causality argument from trade barriers to the effect of comparative advantage on yield

In the paper I show AY's causality from newly imposed trade barriers to the effect of original yield (technology) on labor intensity to be illogical. In a footnote, I refer to this appendix for an explicit argument, along the same lines, in the case of the causality from trade barriers to the effect of comparative advantage on yield.

AY's estimation results of *yield* regressed on factor inputs and weather interacted with time imply that provinces with a comparative advantage (better weather) make ever less use of it in terms of producing yield. But increasing trade barriers cannot account for this finding.

Assume a province has a high comparative advantage value. It then makes ever less use of this comparative advantage in producing yield. First, assume this province is a grain deficit province. Why would an increase in trade barriers cause this province to reduce *yield*, i.e., to waste resources (given that other factor inputs are controlled for)? If anything, an increase in trade barriers should cause a province with a grain deficit to increase either the yield or the sown area of grain, or both.

Second, assume this province is a grain surplus province. Why would an increase in trade barriers cause this province to reduce *yield*? If anything, an increase in trade barriers should cause a province with a grain surplus to reduce the *sown area of grain* (and to use the land for the production of agricultural products in which there is a deficit), without necessary effect on grain *yield*. But suppose a province has a surplus in *all* agricultural products, and suppose farmers must sow all available land at all possible times (an unrealistic assumption), so that the only way for farmers to respond to trade barriers is to work *less enthusiastically*, or to burn their output, possibly across the board, and including on fields sown with grain. (Working *less* is captured by a control variable.)

Then the point remains, as made in the paper for the case of labor intensity (rather than yield) as dependent variable, that a high weather value (high comparative advantage value) *does not imply*, nor is correlated with, provincial grain surplus. In other words, an increase in trade barriers may lead to a change in the degree of provincial grain surplus and possibly a change in yield. But weather can't substitute for provincial grain surplus in this causal chain: it neither (logically) stands for, nor proxies, provincial grain surplus.

Measurement of household grain consumption
(of relevance to agricultural surplus/ deficit argument in paper)

In the paper, I argue that if trade barriers are imposed, what matters for movements in labor intensity is if the province is initially an agricultural surplus or deficit province; initial yield is of no relevance. In footnotes in the paper, I report that initial yield or weather are not even by chance *correlated* with provincial agricultural surplus or deficit. (Therefore, AY's argument does not even hold via a chance correlation.)

One measure of surplus or deficit used in the footnotes in the paper is the ratio of aggregate provincial household grain consumption to provincial grain production. This ratio ignores grain for feed or for industrial use (on which data are not readily available); it is below unity across all provinces.

Provincial household grain consumption is obtained in a round-about way. The volume of grain consumed by households is available as the quantity of urban *nationwide* average per capita grain consumption (*Statistical Yearbook 1986*, p. 679, and *1998*, p. 327, with province-specific data not available) and as the quantity of rural *provincial* average per capita grain consumption (*Statistical Yearbook 1986*, p. 669, and *1998*, p. 351). Urban and rural population data for 1985 are from the *Statistical Yearbook 1986*, p. 93. For 1997 these data are not available, but can be derived from provincial urban and, separately, rural consumption in the national income accounts divided by the published provincial urban and rural per capita consumption values in the national income accounts (*Statistical Yearbook 1998*, pp. 71f). In 1997, for each province, the derived urban and rural population data provide the weights to apply to the (national) urban and (provincial) rural quantity data to obtain a provincial urban-rural average quantity; this average quantity is then multiplied by the official provincial population data (*Statistical Yearbook 1998*, p. 113; not in all provinces do the derived urban and rural population values add up to the official total population value). Provincial sown area of grain and grain production are from the *Statistical Yearbook 1985*, pp. 176 and 182, and from *1998*, pp. 400, and 403.

Agricultural labor in non-crop agricultural production

One alternative explanation for AY's evidence, briefly mentioned in the paper, is a gradual move from a distorted economy to a more market-oriented economy. Assume that in 1978 local self-sufficiency objectives dictated the planting of grain for food even if that was against comparative advantages, and that grain production and grain reserves took priority over all other agricultural products. In 1978, then, agricultural labor focused on grain production, and the yield was determined by soil quality and weather. By 1997, provincial self-sufficiency in grain may no longer have been as important, and provinces, as much as possible given remaining grain quotas, switched to the production of products in which they had a comparative advantage. The coefficient of variation of the share of grain in total sown area increased by one-half from 0.0866 in 1978 to 0.1243 in 1997.

A province with a high value of the comparative advantage variable, may, for example, find it more profitable to switch out of grain production into the production of other agricultural products; grain yield is unlikely to be affected much and any change in labor intensity, defined as *primary sector employment* per sown area, is determined by the labor requirements of other agricultural products in comparison to grain. On the other hand, a comparative disadvantage may mean that agricultural products other than the original staples cannot be produced profitably, and these provinces then stay with the original grain production, perhaps trying to improve yield by improving seed quality or increasing labor intensity.[1] This can explain the 1985 vs. 1997 labor intensity vs. yield patterns and the 1985 yield vs. 1985-97 change in labor intensity pattern shown in the figures in the paper (as well as the regression results).

Econometric issues in Alwyn Young's regressions

(i) AY assumes that yield and factor inputs follow a deterministic trend, which is not particularly likely if seed technology or improvements in machinery are random.

(ii) When regressing the yield on the various factor inputs and the three precipitation measures in equation (1), AY includes labor as independent variable, even though he previously argued that the causality runs from yield to labor; i.e., labor is endogenous.

(iii) AY derives the weather values based on the coefficients of the three measures of precipitation in equation (1), an equation which controls for factor inputs. This would not matter if the precipitation measures and various factor inputs were simply correlated; the coefficients would still be unbiased, just less reliably estimated. But the case here is not just one of multicollinearity, because the precipitation measures affect yields through two channels, namely directly as well as *through* each of the factor inputs.[2] Including the factor inputs in the regression biases the coefficients of the precipitation measures. Biased estimates of the coefficients of the precipitation measures then yield a biased estimate of "weather." (Weather is the aggregation of the three precipitation measures using the coefficients obtained in equation (1), averaged over time for each province.) A biased estimate of weather implies a biased estimate of the impact of weather, i.e., of the comparative advantage of grain production in a particular province, on factor inputs.

(iv) The coefficients in equation (8) appear nearly un-interpretable. The weather variable was constructed following equation (1) and incorporates the various biases of the coefficients of the precipitation measures. This province-specific summary of precipitation, with a trend added, now competes a second time against the factor inputs in explaining yield; it again affects yield not only directly, but also through the factor inputs (as a precipitation summary and as a trend). Furthermore, by reducing three precipitation measures for one province into one with an added deterministic time trend, the scope for explaining variation in yield is reduced, which then gives higher significance levels to the factor inputs.

---

[1] With comparative advantage measured either in form of original yield, as in AY's charts, or using AY's (original) weather variable, as in AY's regressions, this pattern is confirmed. Grain yield (or weather) and share of grain in sown area are (both) negatively and significantly correlated in 1978 and in 1997, at the 0.1% and the 5% significance level.

[2] At least this is what AY's equations (2)-(5) imply. AY regresses the factor inputs on the province-specific, over time constant measure "weather" (interacted with time). For each province, "weather" is the weighted aggregation of the three precipitation measures, averaged over time.

(v) Beyond these technical issues, AY throughout his regressions does not allow province-specific factors to vary over time. But provincial agricultural policies changed substantially over time. All such province-specific, time-dependent factors, if they are somewhat correlated with AY's measure of comparative advantage, end up being captured by the coefficient of his weather variable (one provincial value independent of time) interacted with time. His comparative advantage variable interacted with time, thus, captures not only climate advantages. AY's weather values tend to be low for the most conservative and least-developed provinces (the five provinces with the lowest weather values are Ningxia, Gansu, Inner Mongolia, Tibet, and Xinjiang), although the pattern is not perfectly unambiguous at the top (with the highest weather values for Hunan, Zhejiang, Shanghai, Anhui, and Hubei).

Problems of variables used in Alwyn Young's regressions

(i) Grain and labor are mismatched. His yield measure is a measure of grain output per sown (rather than harvested) grain area, which is then, lacking better data, related to *total employment* in the primary sector relative to total sown area. Grain accounted for 80.34% of all sown area in 1978, for 75.78% in 1985, and for 73.33% in 1997. Furthermore, output of pork, beef, and mutton (also produced by the primary sector employment) in 1997 was 2.32 times 1985 output.[3] If more labor per sown area is required for non-grain production than for grain production, which is true at least for vegetable, meat, and fruit production, then provinces that move out of grain production, perhaps to make better use of comparative advantages in a market economy, experience higher labor intensity in terms of primary sector employment per ha than before.

(ii) The quality of the labor data themselves is questionable. The agriculture section of the *Statistical Yearbook* reports a number of rural laborers in farming, forestry, animal husbandry, and fishery (i.e., in agriculture) that is different from the primary sector employment figure reported in the employment section of the *Statistical Yearbook*; AY uses primary sector employment data from the provincial statistical yearbooks (p. 1132). In 1978, employment in agriculture was 0.49% larger than employment in the primary sector, in 1985 2.50% *smaller*, and in 1997 6.21% smaller (*Statistical Yearbook 2005*, pp. 118, 446.) In what appear the report form employment data in the employment section of the *Statistical Yearbook*, farming, forestry, animal husbandry and fishery in 1978 and 1985 were identical to primary sector employment, but in 1997 5.01% smaller (*Statistical Yearbook 2005*, pp. 118, 125). In other words, AY's employment data may increasingly overestimate employment in agriculture, which would tend to bias his labor coefficient downward in the regression of yield on, among others, labor. (Also see the appendix on labor data.)

(iii) AY's measure of irrigation is calculated as irrigated area divided by total sown area; the division should be by cultivated area, because sown area reflects multiple harvests during one year but irrigated, like cultivated area, does not. While total sown area increased over time, cultivated area declined.

[3] For the grain data see *Rural Yearbook 2005*, p. 129, and for the pork, beef and mutton data *Rural Yearbook 1988*, p. 100, and *Statistical Yearbook 1998*, p. 410; milk output tripled, sheep wool production almost tripled, and poultry egg production almost quadrupled. AY (p. 1120, note 37) is aware of the decreasing share of grain in total area sown, but judges the decline to be too small to account for his results. He does not consider livestock.

(iv) AY's kw power measure covers tractors (28.14% economy-wide in 1997) and diesel engines (10.22%), but also trucks for agricultural use (15.65%), which may not have any impact on yield, motorized fishing boats (2.57%), and a very large implicit residual (43.42%).[4] Presumably only tractors help in determining the yield (and some of the residual if it were to cover other machinery used on fields on which grain is grown).

Correction of four data points in Alwyn Young's dataset, and replacement of two variables

AY's grain output of Sichuan includes Chongqing in all years, but not in 1997; his Sichuan area of grain sown includes Chongqing in all years. The Sichuan/Chongqing grain yield of 1997, thus, is inconsistent. In detail, AY's grain output value for Sichuan plus Chongqing together in 1995 is 4365 (times 10,000 tons), in 1996 4495.7, but in 1997 3461.3. The *Statistical Yearbook 1998* gives a 1997 value for Sichuan alone of 3461.3, and for Chongqing of 1157.7, together 4619. The latter is the value with which I replaced AY's combined Sichuan/Chongqing 1997 value.

AY's grain output of Guangdong/Hainan in 1997 exceeds the combined values as published in the *Statistical Yearbook 1998*, p. 403, by 54.83%. In detail, AY's grain output value for Guangdong plus Hainan together in 1995 is 1936.6, in 1996 2036.9, and in 1997 3269.3. The *Statistical Yearbook 1998* gives a 1997 value for Guangdong of 1897.7 and for Hainan of 213.9, together 2111.6. The latter is the value with which I replaced AY's combined Guangdong/Hainan 1997 value. It happens to be 1157.7 smaller than AY's wrong value, a difference equal to Chongqing's 1997 grain value. (It looks like Chongqing's 1997 grain output was mistakenly added to Guangdong/Hainan rather than to Sichuan.)

AY's primary sector employment value for Shandong in 1995 is 2515 (times 10,000 persons), in 1996 3919.8, and in 1997 2487.9. The values in the *Statistical Yearbook* series are 2515.7, 2487.9, and 2510.5. It looks like the *Statistical Yearbook* value of 1996 ended up in the 1997 position in AY's dataset, while the primary sector employment 1996 value in AY's dataset is clearly out of line. I replaced the 1996 and 1997 Shandong primary sector employment values in AY's dataset by the *Statistical Yearbook* values.

As explained in the previous section, AY's measure of irrigation relates the area of irrigated fields to the area sown rather than the area cultivated. The distinction between sown and cultivated is that cultivated land is land under cultivation, while sown land double-counts that cultivated land which is sown twice a year, and triple-counts that cultivated land which is sown thrice a year (but the measure "area under irrigation" does not). I therefore divide area under irrigation by cultivated area. For 1996 and 1997, no data on cultivated area are available, and I use the 1995 values. While not perfect, this seems preferable to a logically flawed measure (that relates irrigated to sown area). The data are from the *Statistical Yearbook* series.

Also as explained in the previous section, AY's measure of agricultural machinery used on fields sown with grain, namely total power in agriculture, is by far too comprehensive. Four

[4] For the data see *Statistical Yearbook 1998*, pp. 391f.

alternatives are the capacity (in kw) of large and medium-sized tractors, the capacity of small tractors, the number of units of large and medium tractor towing farm machinery, and the number of units of small tractor towing farm machinery. The first and third differ little over time and fare poorly in AY's regressions. The second and fourth are both improvements over AY's measure, and the second comes closest to what AY seemed to want to measure (and is used to replace AY's power measure). The data are from the *Statistical Yearbook* series.

The correction of AY's typos and the replacement of the two variables improves most of his results, including all of his (desired) core results. An insubstantial change is that small tractors are not significant in his final regression of yield on weather times time (but are significant in his first regression of yield on measures of precipitation, in contrast to his power variable).

Some of the data on Tibet in 1992 and 1993 are identical, in AY's dataset as well as in the *Statistical Yearbook*: the data for output of grain (and fruit or meat), sown, cultivated, and irrigated area, power/ agricultural machinery, and fertilizer. I proceed as AY and simply ignore the problem. No data on agricultural machinery/ power for Tibet in 1997 are available and the mean of the 1996 and 1998 values is used here. Tibet data, in general, appear rather poor when year-to-year comparisons are made, with changes that are not always plausible. Occasionally dropping Tibet from a regression makes no big difference, so I retained it, as did AY. (I have not systematically rerun *all* regressions without Tibet.)

Explanation of (other) variables used

Using data from the *Statistical Yearbook* series I obtain (and use) the same yield measure as AY (after correcting for the typos in his series). I use AY's labor data (with the correction of the two typos) although other, and often slightly different, series from the *Statistical Yearbook* or *Fifty Years* are available; switching in some regressions to alternative labor series makes some but no major or systematic difference. My fertilizer data (from the *Statistical Yearbook* series) are identical to AY's. Whenever I use the same data or variable as AY, I use the same units as he does (which often is 10 times what he says it is, in the variable explanations under his table with regression results).

My precipitation data (from the *Statistical Yearbook series*) are identical to AY's except in four instances (four provinces in a particular year). In seven instances when the sum of monthly precipitation does not add up to the annual figure, I use the sum of monthly precipitation to obtain average monthly precipitation because the coefficient of variation inevitably uses a standard deviation that is based on the monthly data. In three of these seven instances, AY appears to use the annual figure to obtain mean monthly precipitation, but then has no choice but to use the monthly data to obtain the standard deviation. AY's 1997 precipitation data for Sichuan/Chongqing are those of Chongqing only; I use the combined Sichuan/Chongqing data (as imposed by the data for all other years and as done by AY in 1997 for other variables). I follow AY's practice of assuming zero rainfall when the *Statistical Yearbook* for a particular province in a particular month has no entry (plausible in cross-period comparisons). When I construct a measure of weather to reproduce AY's results I use AY's precipitation data. When I use all nine climate measures to construct a measure of weather, I use my precipitation data.

Temperature and sunshine values are also based on the monthly data (in the *Statistical Yearbook series*), which do not always match the published annual values (6 instances of sunshine in 1985 and 1988, very minor instances of temperature discrepancies). Temperature data are incomplete for Tibet in 1992 and 1993, and sunshine data for Shanghai in 1986 and Tibet in 1985 and 1986 (the climate measures for these provinces in these years then are omitted).

In the variable "fruit-grain-benefit"—i.e., in the variable ln(net benefit [in yuan RMB] per laborer of growing fruit/ of growing grain)—net benefit (*jianshui chun shouyi*) denotes total product value less total product costs less tax, and laborers are measured in standardized labor days. The data, only available at the national level, are from the *Rural Yearbook* series. 1990 data are not available; 1990 laborers are obtained as the mean of 1989 and 1991 values, and 1990 net benefits are interpolated between 1989 and 1991 values using the 1989, 1990, and 1991 prices of related agricultural products from the *Price Yearbook*. (For example, if the price of a unit of wheat in 1989 was 1 yuan, in 1990 5 yuan, and in 1991 6 yuan, the 1989 benefit value received 80% weight and the 1991 benefit value 20% weight in calculating the 1990 benefit value.)

Data sources of other variables are listed in the following; individual data points that may be missing in the *Statistical Yearbook* or in *Fifty Years* are supplemented wherever possible from *Seventeen Years* and the *Rural Yearbook*. (The list includes information on variables used only in additional regressions presented in this appendix.)

Fertilizer: *Statistical Yearbook* (total across all chemical fertilizers listed, as is AY's practice).

Product output measures: grain (*liangshi*), fruit (*shuiguo*), and meat (pork, beef, and mutton (*zhuniuyangrou*)): *Statistical Yearbook*.

Relative productivity: employment data from *Fifty Years*, value added from *GDP 1952-95* and *GDP 1996-2002*.

SOE share in investment: *Statistical Yearbook*.

Agricultural share in rural labor: agriculture section of *Statistical Yearbook* with employment in rural areas (*xiangcun*) and in agriculture (*nonglinmufuyuye* prior to 1993, *nonglinmuyuye* since 1993).

Rural and urban household income: *Fifty Years*.

Rural employment: *Fifty Years* for both employment as well as the total employment used in this instance (for both data points to come from one, possibly consistent source).

Area sown with particular product: *Statistical Yearbook*.

Farm output (from the *Statistical Yearbook* series), a variable used only in the regressions presented in this appendix, reflects the share of farming in the gross output value of agriculture. Gross output value of *agriculture* covers *nongye* (agriculture) prior to 1993 and *nonglinmuyuye* (the exhaustive categories farming, forestry, husbandry, and fishery) since 1993. This means that the gross output value of agriculture since 1993 excludes *fuye* (sideline activities). The national aggregate data published in and after 1993 for earlier years have the same values as the pre-1993 data published prior to 1993. Sideline activities tended to account for about 3-6% of gross output value of agriculture in 1978-93.

The national gross output value of *farming* published since 1993 for earlier years exceeds that published earlier. The pre-1993 national gross output value of farming *plus* sideline activities exceeds the later (post-1992) published gross output value of farming of pre-1993 in 1985 through 1992 by 0.06 to 0.08%. This suggests that virtually all sideline activities are since 1993 regarded as farming activities. Using the proportion of the national data each year, i.e., the national fraction of sideline activities that is later counted as farming activities (almost all), this proportion is applied to the provincial sideline activity data in order to construct provincial gross output value of farming for the years prior to 1993 that match the later definition. The data are taken from the *Statistical Yearbook* series; *Fifty Years* would have had similar data, except that some provinces appear to report data for all years according to the old definition, while others do so according to the new definition (and Beijing reports "real" values, and a few other provinces' time series data are not credible).

Gross output value was re-defined in 1995 to newly exclude value added taxes. The change is explicit in the industry data of the *Statistical Yearbook*, but not in the agricultural gross output value data. Because it affects both agriculture and farming, and presumably equally, the ratio should still be consistent.

In the regressions, I have toyed with the relative shares of different farm products in sown area, and with a variety of relative benefit measures. Across a number of regressions, the same set of measures tended to be significant. The one of each used in the regression reported in the paper also appears to make most sense because they cover the major categories with the bigger differences. The first regression in Table 2 below includes all measures of each set that tended to be somewhat significant.

The data cover 29 provinces over the period 1985-97 (AY's period). Hainan data separate from Guangdong data became available starting 1988 and Chongqing data separate from Sichuan starting 1997; these provinces in the following years were folded back into Guangdong/Hainan and Sichuan/Chongqing.[5] Because Qinghai has no values for the area sown with rice, Qinghai drops out of the regressions whenever the variable "rice-grain-sown" (ln(sown area of rice/ sown area of grain)) is included. It is possible that Qinghai does not grow rice, but I am not perfectly sure. (I have alternatively set Qinghai's values of this variable equal to zero, which does not change the results of the regressions, but I have not done this systematically across all regressions reported in the paper and in this appendix).

Yet more regressions, and interpretation of coefficients of control variables

---

[5] *Fifty Years* has a complete Hainan series for all years since 1952, and Chongqing data since 1996; its Sichuan data in all years cover only Sichuan. *Fifty Years* does not cover all variables.

Table 1 in the first three columns copies AY's original results. For contrast, columns four to six report the results with corrected values and two replaced variables, but otherwise the same setup as AY's (the same data are reported in the paper).

The last two columns of Table 1 switch to the more comprehensive weather variable (based on all 9 climate variables) and include two control variables to control for the fact that employment in the primary sector is not all preoccupied with grain production. In contrast to the results reported in the fifth column, labor now moves *with*, rather than against comparative advantage. (This result is even stronger if irrigation, fertilizer, and small tractors are omitted, as by AY in his regression.) I.e., the results in AY's labor regression appear confounded by the fact that his labor variable relates *all* employment in the primary sector to sown area, irrespective of what these agricultural laborers are actually doing. The last regression reported in Table 1, with yield as the dependent variable, comes with similar findings as AY's, except that the significance levels of the key independent variables (labor, comparative advantage) are now much lower.

Table 2 and Table 3 present a number of variations of the 3SLS regressions. Table 2 starts with the largest number of control variables. The key findings are the same as reported in the paper. In terms of control variables, with labor intensity (number of primary sector laborers per area of sown land) as dependent variable:

- the more fruit output relative to grain (measured in weight), the higher labor intensity—fruit production is more labor intensive than grain production (fruit production requires about five times more labor per land area than grain),
- the higher the value added per laborer in the primary sector relative to provincial GDP per laborer, the lower is labor intensity—relatively high income comes with extensive farming,
- the higher the SOE share in investment expenditures, the higher labor intensity—a province in which the state sector is strong is likely to offer little informal employment and laborers are stuck in agriculture,
- the larger the share of agricultural laborers in the rural labor force, the higher labor intensity—the fewer non-agricultural jobs can be found in the countryside, the more likely rural laborers end up remaining in agriculture,
- the higher rural relative to urban income, the lower labor intensity—this is similar to the labor productivity argument, but here in terms of household income,
- the larger meat relative to grain output (measured in weight), the higher labor intensity—meat production is more labor intensive than grain production,
- the higher gross output value of farming relative to that of agriculture, the lower labor intensity—this is again similar to the labor productivity argument, but here in gross output terms and with farming related to all of agriculture rather than agriculture to non-agriculture,
- the higher the share of rural laborers in the economy-wide labor force, the higher labor intensity—this is similar to the argument based on the share of agricultural laborers in the rural labor force;

The effects of the control variables on the dependent variable grain yield (grain output per sown area) are:

- the larger the share of rice in area sown, the higher is yield—rice is a high-yield grain (which can be confirmed by comparing grain-specific sown areas to the corresponding output; rice has an approximately 50 percent higher yield per land area than wheat),
- the larger the share of corn in area sown, the lower is yield—corn is an average-yield grain, and the correlation is barely significant,
- the larger the share of tubers in area sown, the higher is yield—tubers, in the period under consideration, is an average-yield grain (converted by China's National Bureau of Statistics into "grain" at a ratio of 5:1, i.e., 5kg of tubers are counted as 1kg of grain) and the relatively strong positive correlation is unexpected,
- the larger the net benefit per laborer of growing fruit rather than grain, the higher the (grain) yield—if net benefits outside grain production are high, grain production is likely to be minimized, within externally given constraints, to where it is most productive, or land sown with grain is worked carefully to meet grain output targets while pursuing other agricultural production.

The next three sets of 3SLS results reported in Table 2 (ii-iv) and the first set of 3SLS results reported in Table 3 (v) first reduce the number of variables in the labor regression, then, over three scenarios, the number of variables in both equations down to the bare minimum. The sign and significance of the coefficients of weather interacted with time in the three scenarios don't change (and the size of the coefficients stays almost constant) through these variations, i.e., comparative advantage has a positive impact on labor intensity, and none on yield. The only finding that changes is that labor in the yield regression in the last version reported in Table 2 turns (just barely) insignificant.

Table 3 then continues in the 6th through 8th scenario with the 3SLS results reported in the paper but adds two further variables interacted with time, first each individually, and then together. The reason to add further variables interacted with time is to explore to what extent the comparative advantage variable used so far, weather (based on nine climate variables), loses importance when other measures of comparative advantage are included. To provide the answer right away: while the weather effect becomes perhaps slightly weaker, its positive significance for labor intensity and its irrelevance for yield do not change.

The first variable to be added is the 1985 share of SOEs in investment expenditures interacted with time. One rationale would be to view the SOE share as a negative measure of the degree of reform-mindedness of a province: the larger the SOE share in 1985, the less reform-minded is the province. An opposite rationale would be that the SOE share in 1985 proxies for the degree of original industrial infrastructure or the basis from which private entrepreneurship can take off; the larger the role of SOEs early on, the better the industrial foundation. The findings—yield is positively associated with the 1985 SOE share interacted with time, and labor intensity in agriculture negatively—would suggest the latter interpretation; a sound industrial foundation has a positive effect on agricultural yield while drawing labor out of agriculture.

The second new variable, also interacted with time, is the 1985-95 average ratio of sown to cultivated area as a measure of how suitable the climate is for agriculture.[6] The more harvests

[6] The data on cultivated area only go through 1995, as mentioned earlier, therefore the choice of 1985-95 rather than through 1997.

per year, the higher the ratio. The ratio has a negative impact on yield over time, perhaps a sign that agricultural land becomes exhausted more rapidly if it is sown more frequently in a given time period. The ratio has no impact on labor intensity. In the final regression with both new variables included, the results are similar (with now also a barely significant negative impact on labor intensity).

Yet another variation, not reported in the tables, is to use the climate information available for 1951-80, with monthly provincial averages *across all forty years*. These nine, *historic*, province-specific climate variables (as in the fourth column of the regression table in the paper) now are constant over time, i.e., directly represent comparative advantages. Regressing yield as in AY's regression (third column) with, instead of AY's weather variable, the nine climate variables (measuring provincial characteristics, and precluding the use of provincial dummies) as well as the nine climate variables *interacted with time* (and including the same controls as AY), only one interaction with time is significant (temperature squared); if the number of climate (comparative advantage) interaction variables is reduced to three, namely average monthly temperature, precipitation, and sunshine (or, alternatively the coefficient of variation across months), only sunshine (alternatively, the coefficient of variation of precipitation) interacted with time has a significant impact on yield (positive).[7]

U.S.

Data sources

Labor is total farm employment from http://www.bea.doc.gov/bea/regional/spi/default.cfm, accessed on 16 December 2005 for 1977-2001 data by SIC industry, and accessed 22 December 2005 for 2001-2004 data by NAICS industry (two different classifications for the two periods); the farm employment values of 2001 in the two sources are identical.[8] Data on crop output (either volume or weight measure) and harvested area (in acres) are available from the agricultural censuses, at http://151.121.3.33:8080/Census/Create_Census_US.jsp, accessed 22 December 2005, covering the approximately 5-yearly agricultural censuses with the most recent in 2002.[9]

In the case of China, the land data refer to area *sown*, not harvested; the U.S. agricultural census does not include data on grain-specific area sown. The U.S. agricultural census data do not cover Alaska, D.C., and Hawaii. The data for 1978, in contrast to 2002, come with the label

---

[7] With labor intensity as dependent variable, and year, irrigation, fertilizer, and small tractors as controls, 8 of the 9 interactions terms are significant (3 negatively). Using only 3 interaction terms, temperature and sunshine have a positive effect and precipitation a negative effect; when the coefficient of variation is used, the signs are exactly the opposite.

[8] Either there has been a change in data availability, or my earlier source information (as presented in this paragraph) contains a typo. As of 30 January 2006, the website provides *1958*-2001 data by SIC industry; this time coverage (going back further) appears new. The original choice of 1978 as earliest year in the analysis here was dictated by the availability of employment data (starting in 1977); otherwise the earliest agricultural census data available, of 1974, could have been used.

[9] The years with agricultural censuses for which data are available online are 1974, 1978, 1982, 1987, 1992, 1997, and 2002.

"not adjusted for coverage." Examining the 1997 data, the only ones available without and with this label, there is no difference in the data. (Perhaps the label is relevant for non-grain, or non-crop variables.) 1997 data were also compared to 1978 and 2002 data. The 1997 and 2002 data are more extensive than those of earlier censuses. The patterns reported below for 2002 are highly similar in 1997.

Grain measured in bushels

The figure in the paper reports the results with a cut-off point for labor intensity of 0.1. This implies the omission of, in 1978, Arizona (labor intensity of 0.17), California (0.18), Connecticut (3.42), Florida (0.15), Maine (0.33), Massachusetts (3.18), Nevada (0.13), New Hampshire and Rhode Island (both with no crop output and acreage data in the agricultural census), Vermont (1.42), and West Virginia (0.33); and in 2002 of Alabama (0.13), Arizona (0.13) California (0.46), Connecticut (3.34), Florida (2.03), Georgia (0.11), Maine (0.20), Massachusetts (3.94), Nevada (0.42), New Hampshire (5.81), Rhode Island (17.22), Utah (0.12), Vermont (1.23), and West Virginia (0.42).

High values presumably simply reflect the fact that some states grow very few crops measured in bushels—agriculture in Florida, for example, may specialize in the production of citrus fruits—and the total agricultural labor force is then related to the few acres on which crops measured in bushels are grown. The same problem obviously also besets Chinese data. However, specialization in China is unlikely to have gone as far as in the U.S., given the pre-reform imperative of (and even during the reform period stress on) local self-sufficiency in grain production. In any case, the decision to relate *total* agricultural employment per sown area to *grain* yield is AY's in the first place.

Maintaining the 0.1 level of labor intensity as cut-off point, but switching to natural logarithms, to exactly match AY's original charts, the slopes are weaker but the Chinese pattern of change continues to hold for the U.S., too (Figure 1).

Figure 2 reports the findings with a cut-off point for labor intensity of 0.3, and using the values as they are, i.e., not in natural logarithms. Compared to Figure 1, it contains four more observations in 1978, and five more observations in 2002. Both trend lines, of 1978 and 2002, are now slightly upward sloping. However, (i) the slope is so small that it appears not significantly different from zero in either year, and (ii) the figure clearly shows how the far outliers impact on the slope of the trend line. For example, dropping Maine in 2002 leads to a slightly negative slope of the 2002 trend line.

Figure 3 shows the same data in natural logarithms, following AY's presentation of the Chinese data. The 1978 trend line continues to have a slightly positive slope, but the 2002 trend line now has a faintly negative slope. If one were to view these slopes as significant, the U.S. pattern again matches that of China.

All grains

The agricultural censuses cover grains measured in bushels—corn for grain, wheat for grain, oats for grain, barley for grain, sorghum for grain, and soybeans for beans—but also corn for silage or greenchop, sorghum for silage or greenchop, dry edible beans, potatoes, sweet potatoes, and rice, all measured in tons or cwt (hundredweight). The agricultural census data of 1978 for grains not measured in bushels appear exceedingly incomplete. For example, the categories corn and sorghum for silage or greenchop, dry edible beans, and sweet potatoes carry no entry in any state. Grains measured in bushels accounted for 100% of grain output (all items) in about two-thirds of all states. In contrast, in 2002 all states carried data entries for grains not measured in bushels.[10] This would suggest focusing on grains measured in bushels for the comparison of 1978 and 2002, as presented in the paper and above.

Proceeding, nevertheless, with the available data requires a conversion of the different output measures into one output measure. Ideally, for comparability, this follows the Chinese practice. The Chinese grain data cover paddy rice, wheat, corn, sorghum, millet, "miscellaneous grains," beans, and tubers (potatoes and sweet potatoes, but not taro and cassava); tubers are turned into "grain" at the rate of 5:1, i.e., 5kg potatoes are counted as 1kg grain (*Statistical Yearbook 2005*, p. 480). For the U.S., all crop data that match the Chinese definition are used. These include, now with the measurement they come in: corn for grain (bushels), corn for silage or greenchop (tons) [only for corn and sorghum a separate category "for silage or greenchop" is available], wheat for grain (bushels), oats for grain (bushels), barley for grain (bushels), sorghum for grain (bushels), sorghum for silage or greenchop (tons), soybeans for beans (bushels), dry edible beans excluding lima beans (cwt), potatoes (cwt), sweet potatoes (cwt), and rice (cwt). One bushel is 35.24 liters and was translated 1:1 from liter into kilogram; one hundredweight (cwt) is 100 pounds, which in turn is 45.359kg.

Figure 4 reports the results if all observations with a labor intensity higher than 0.3 are removed. The omitted observations—different from the case when only those grains measured in bushels are covered, because the total number of agricultural laborers is now related to the area on which the more comprehensively defined grains are harvested—in 1978 are Connecticut (labor intensity of 3.42), Massachusetts (3.18), New Hampshire (14.40), Rhode Island (0.50), Vermont (1.32), and West Virginia (0.33); in 2002 they are Connecticut (0.34), Florida (0.73), Massachusetts (0.47), New Hampshire (0.34), Rhode Island (0.49), and West Virginia (0.31); in addition, Vermont was excluded in 2002 because its yield was twice the level of any other included state.

The figure, for the U.S., in comparison to China shows an inverse development over time, from, following AY's logic, perfect trade barriers in 1978 to less severe trade barriers in 2002. Omitting the right half of the figure, i.e., those states with a labor intensity above 0.1, yields the same pattern of trend lines. Switching to natural logarithms (with either cut-off point for labor intensity) also yields the same pattern of trend lines. (Given that there is no difference in pattern, these additional figures are not included here.)

---

[10] In part this could be an artifact in that in 1978 no state had any entry for "corn for silage or greenchop," a non-bushel item, and this output was possibly included in "corn."

Table 1. Productivity, Factor Allocations, and the Weather

	AY's original results			Typos corrected, 2 variables replaced				
	Yield	Labor	Yield	Yield	Labor	Yield	Labor	Yield
Year	.009**	.006**	.026**	.003	.006**	.019**	-.027**	.019**
	(3.1)	(3.7)	(5.9)	(1.1)	(3.4)	(4.7)	(-4.9)	(3.0)
Irrigation-AY	.309**		.199**					
	(6.3)		(4.5)					
Irrigation				.352**		.271**	.024	.250**
				(7.6)		(6.6)	(.6)	(5.6)
Fertilizer	.214**		.155**	.202**		.163**	-.038	.189**
	(6.2)		(4.3)	(6.5)		(5.0)	(-1.3)	(5.7)
Power	-.043		-.080*					
	(1.0)		(1.9)					
Small tractors				.044**		.024	.038*	.038*
				(2.6)		(1.4)	(2.5)	(2.2)
Labor	-.091		-.148*	-.122*		-.166**		-.132*
	(1.5)		(2.4)	(-2.12)		(-2.9)		(-2.1)
Precipitation	.021**			.022**				
	(2.6)			(3.1)				
Precipitation squared	-.001**			-.001**				
	(2.7)			(-3.4)				
Coeff. of var. of prec.	-.040*			-.035*				
	(1.9)			(-1.9)				
Weather-AY*time		-.155**	-.260**					
		(4.4)	(5.6)					
Weather-AY (corrected) * time					-.157**	-.253**		
					(-4.1)	(-5.3)		
Weather*time							.089**	-.031*
							(5.8)	(-1.7)
Rice-grain-sown							.042	.118**
							(1.6)	(4.0)
Fruit-grain-benefit							.039**	.041**
							(3.12)	(2.9)
Prov. dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R^2	.942	.975	.944	.954	.978	.954	.981	.954
Number of obs.	374	377	376	375	377	377	364	364

* / ** : significant at 10% / 1% level. T-statistics are in parentheses.

Yield: ln(grain output (10 tons)/ grain sown area (ha)). (The choice of "10" units follows the practice of AY.) AY's data, without and then with correction of typos.

Irrigation-AY: ln(irrigated area/ total sown area). AY's data.

Irrigation: ln(irrigated area/ cultivated area).

Fertilizer: ln(chemical fertilizer (10 tons)/ total sown area (ha)). AY's data

Power: ln(total power of agricultural machinery (10 kw)/ total sown area (ha)). AY's data.

Small tractors: ln(power of small tractors and hand-led tractors (kw)/ total sown area (ha)).

Labor: ln(employment in primary sector (10 persons)/ total sown area (ha)). AY's data, without and then with correction of typos.

Precipitation: average (for each year) monthly rainfall (cm). AY's data.

Weather-AY: (province-specific) mean (of 1985-97) value of precipitation variables multiplied by their coefficients in regression of yield on precipitation variables.

Weather-AY (corrected): (province-specific) mean (of 1985-97) value of precipitation variables multiplied by their coefficients in regression of yield on precipitation variables, where the four typos are corrected and the irrigation and power measures are replaced by the two more appropriate variables.

Weather: (province-specific) mean (of 1985-97) value of 9 climate variables multiplied by their coefficients in regression of yield on 9 climate variables (reported in paper).

Rice-grain-sown: ln(sown area of rice/ sown area of grain). Since Qinghai province has no data on rice, inclusion of this variable implies the exclusion of Qinghai.

Fruit-grain-benefit: ln(net benefit [in yuan RMB] per laborer of growing fruit / of growing grain); national variable, i.e., not province-specific.

Table 2. Productivity, Factor Allocations, and the Weather: 3SLS

	(i)		(ii)		(iii)		(iv)	
	Labor	Yield	Labor	Yield	Labor	Yield	Labor	Yield
Year	-.030**	.008	-.027**	.009	-.024**	.008	-.014*	.019**
	(-3.2)	(1.0)	(-3.8)	(1.3)	(-2.8)	(1.2)	(-1.9)	(2.9)
Irrigation	-.158*	.275**	-.122*	.275**	-.101	.252**	-.134*	.254**
	(-2.5)	(6.0)	(-2.5)	(6.1)	(-1.6)	(5.7)	(-2.0)	(6.0)
Fertilizer	-.080*	.219**	-.048	.218**	-.097*	.186**	-.140*	.187**
	(-1.7)	(6.2)	(-1.4)	(6.3)	(-2.0)	(5.7)	(-2.5)	(5.9)
Small tractors	.010	.073**	.015	.071**	.014	.047**	.014	.041*
	(.7)	(3.7)	(1.1)	(3.6)	(1.0)	(2.7)	(.9)	(2.4)
Labor		-.586**		-.551**		-.345**		-.182
		(-5.8)		(-5.4)		(-3.6)		(-1.6)
Yield	.436*		.335*		.388*		.369*	
	(2.1)		(2.2)		(1.7)		(1.7)	
Weather*time	.065**	-.003	.065**	-.008	.064**	-.012	.065**	-.027
	(4.2)	(-.2)	(4.5)	(-.4)	(4.1)	(-.6)	(3.8)	(-1.4)
Fruit-grain-output	.050**		.050**		.059**		.052*	
	(2.9)		(3.4)		(2.9)		(2.4)	
Relative productivity	-.211**		-.226**		-.247**			
	(-4.4)		(-5.1)		(-4.8)			
SOE share in investment	.068*		.082**		.066*			
	(2.2)		(3.0)		(2.2)			
Agr. share in rural labor	.849**		.808**		.893**		1.035**	
	(8.6)		(8.6)		(10.1)		(10.4)	
Relative income	-.063							
	(-1.5)							
Meat-grain-output	.048*							
	(1.7)							
Farm output	-.242**		-.276**					
	(-3.6)		(-4.4)					
Rural employment	.761**		.702**					
	(6.5)		(7.0)					
Rice-grain-sown		.131**		.129**		.135**		.113**
		(4.4)		(4.4)		(4.7)		(4.1)
Corn-grain-sown		-.028		-.031*				
		(-1.6)		(-1.8)				
Tubers-grain-sown		.060**		.057**				
		(2.8)		(2.7)				
Fruit-grain-benefit		.063**		.061**				.050**
		(4.3)		(4.1)				(3.8)
Prov. dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R^2	.988	.948	.988	.949	.986	.950	.981	.954
Observations	356	356	356	356	364	364	364	364

* / ** : significant at 10% / 1% level. Z-statistics are in parentheses.

In the last 2 columns, dropping the agricultural share in rural labor, or including the SOE share in investment make no difference for the core coefficients (of weather*time, labor and yield).

Yield: ln(grain output (10 tons)/ grain sown area (ha)). (The choice of “10” units follows the practice of AY.) AY’s data with correction of typos.

Irrigation: ln(irrigated area/ cultivated area).
Fertilizer: ln(chemical fertilizer (10 tons)/ total sown area (ha)). AY's data.
Small tractors: ln(power of small tractors and hand-led tractors (kw)/ total sown area (ha)).
Labor: ln(employment in primary sector (10 persons)/ total sown area (ha)). AY's data with correction of typos.
Weather: (province-specific) mean (of 1985-97) value of 9 climate variables multiplied by their coefficients in regression of yield on 9 climate variables (reported in paper).
Fruit-grain-output: ln(fruit output (tons)/ grain output (tons)).
Relative productivity: ln(value added per laborer in primary sector/ (provincial) GDP per laborer).
SOE share in investment: ln(investment expenditures by state-owned units/ total investment expenditures).
Agr. share in rural labor: ln(agricultural laborers/ rural labor force).
Relative income: ln(per capita rural household net income (*chun shouru*)/ per capita urban household disposable income (*ke zhipei shouru*)).
Meat-grain-output: ln(meat output (tons)/ grain output (tons)).
Farm output: ln(gross output value of farming/ gross output value of agriculture).
Rural employment: ln(employment in rural areas (*xiangcun*)/ total employment).
Rice-grain-sown: ln(sown area of rice/ sown area of grain). Since Qinghai province has no data on rice, inclusion of this variable implies the exclusion of Qinghai.
Corn-grain-sown: ln(sown area of corn/ sown area of grain).
Tubers-grain-sown: ln(sown area of grain/ sown area of grain).
Fruit-grain-benefit: ln(net benefit [in yuan RMB] per laborer of growing fruit / of growing grain); national variable, i.e., not province-specific.

Table 3. Productivity, Factor Allocations, and the Weather: 3SLS continued

	(v)		(vi)		(vii)		(viii)	
	Labor	Yield	Labor	Yield	Labor	Yield	Labor	Yield
Year	-.043**	-.0004*	-.027**	.019**	-.027*	.050**	-.015	.053**
	(-4.9)	(-1.8)	(-3.5)	(2.9)	(-2.2)	(5.8)	(-1.2)	(5.6)
Irrigation	-.075	.260**	-.087*	.229**	-.100*	.167**	-.106*	.165**
	(-.9)	(5.3)	(-1.7)	(5.3)	(-2.0)	(3.7)	(-2.2)	(3.6)
Fertilizer	-.121*	.171**	-.071*	.125**	-.098*	.091**	-.078*	.094**
	(-1.8)	(4.5)	(-1.9)	(3.7)	(-2.6)	(2.6)	(-2.2)	(2.7)
Small tractors	.039*	.061**	.012	.054**	.017	.028	.001	.025
	(2.2)	(2.9)	(.8)	(3.2)	(1.2)	(1.6)	(.1)	(1.4)
Labor		-.678**		-.432**		-.519**		-.516**
		(-3.3)		(-4.5)		(-5.3)		(-5.2)
Yield	.468*		.391*		.416*		.366*	
	(1.7)		(2.2)		(2.1)		(2.0)	
Weather*time	.101**	-.011	.056**	.009	.065**	-.018	.043**	-.022
	(5.1)	(-1.4)	(3.9)	(.5)	(4.0)	(-1.0)	(2.7)	(-1.1)
1985 SOE share in investm. * time			-.011*	.020**			-.018**	-.004
			(-2.3)	(4.1)			(-3.6)	(-.5)
Ave. sown/cultiv. area * time					$-9*10^{-6}$	-.016**	-.007*	-.017**
					(-.0)	(-6.2)	(-2.1)	(-4.6)
Fruit-grain-output	.046*		.059**		.066**		.057**	
	(1.7)		(3.5)		(3.7)		(3.3)	
Relative productivity			-.245**		-.266**		-.260**	
			(-5.7)		(-6.2)		(-6.4)	
SOE share in investment			.044		.064*		.046	
			(1.4)		(2.0)		(1.5)	
Agr. share in rural labor			.948**		.861**		.861**	
			(10.2)		(8.3)		(8.7)	
Rice-grain-sown		.152**		.128**		.105**		.102**
		(4.6)		(4.6)		(3.7)		(3.6)
Fruit-grain-benefit				.054**		.062**		.063**
				(3.9)		(4.5)		(4.5)
Prov. dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R^2	.974	.940	.986	.953	.985	.954	.986	.954
Observations	364	364	364	364	364	364	364	364

* / ** : significant at 10% / 1% level. Z-statistics are in parentheses.

Yield: ln(grain output (10 tons)/ grain sown area (ha)). (The choice of "10" units follows the practice of AY.) AY's data with correction of typos.

Irrigation: ln(irrigated area/ cultivated area).

Fertilizer: ln(chemical fertilizer (10 tons)/ total sown area (ha)). AY's data.

Small tractors: ln(power of small tractors and hand-led tractors (kw)/ total sown area (ha)).

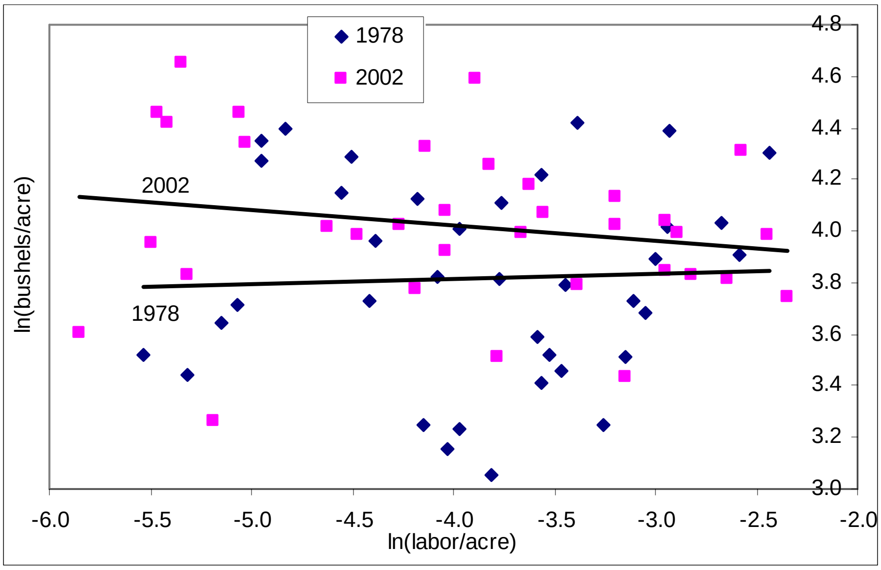
Labor: ln(employment in primary sector (10 persons)/ total sown area (ha)). AY's data with correction of typos.

Weather: (province-specific) mean (of 1985-97) value of 9 climate variables multiplied by their coefficients in regression of yield on 9 climate variables (reported in paper).

1985 SOE share in investm. * time: SOE share in investment, 1985, times time.

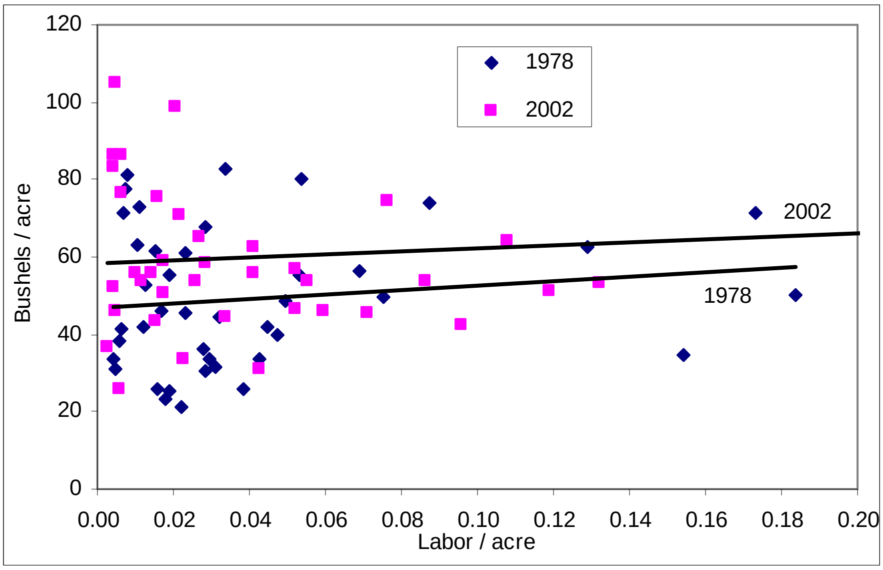
Ave. sown/cultiv. area * time: average 1985-95 'total sown area/ cultivated area, times time.

Fruit-grain-output: ln(fruit output (tons)/ grain output (tons)).
Relative productivity: ln(value added per laborer in primary sector/ (provincial) GDP per laborer).
SOE share in investment: ln(investment expenditures by state-owned units/ total investment expenditures).
Agr. share in rural labor: ln(agricultural laborers/ rural labor force).
Rice-grain-sown: ln(sown area of rice/ sown area of grain). Since Qinghai province has no data on rice, inclusion of this variable implies the exclusion of Qinghai.
Fruit-grain-benefit: ln(net benefit [in yuan RMB] per laborer of growing fruit / of growing grain); national variable, i.e., not province-specific.



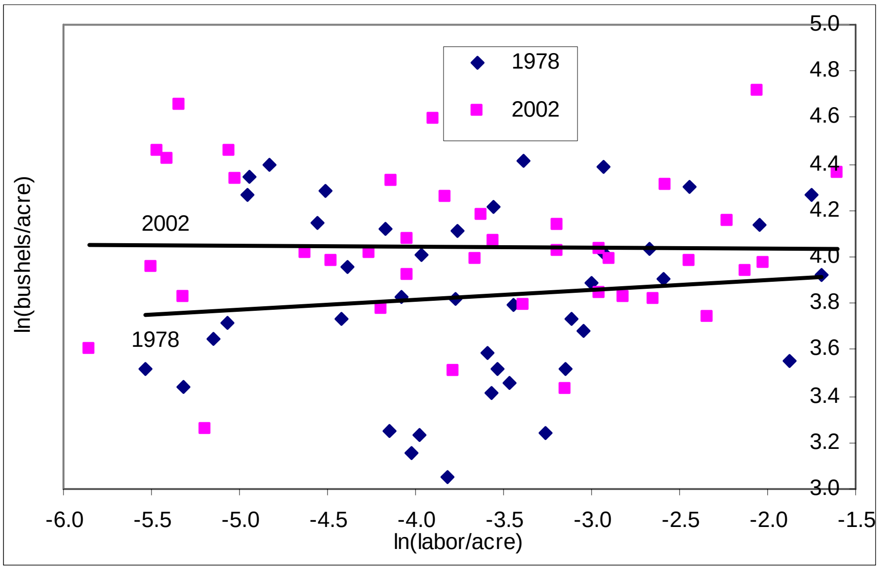
A cut-off point of a labor intensity of 0.1 is applied. Only those grains are included which in the agricultural census are measured in bushels (corn for grain, wheat for grain, oats for grain, barley for grain, sorghum for grain, soybeans for beans).

Figure 1. Factor Intensity and Productivity, Grains Measured in Bushels (U.S. States), in natural logarithms



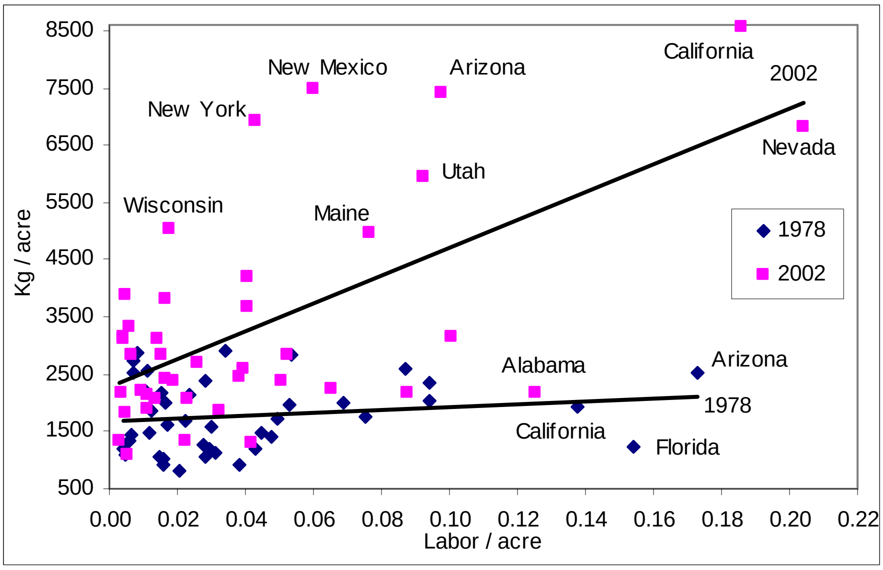
A cut-off point of a labor intensity of 0.3 is applied. Only those grains are included which in the agricultural census are measured in bushels (corn for grain, wheat for grain, oats for grain, barley for grain, sorghum for grain, soybeans for beans).

Figure 2. Factor Intensity and Productivity, Grains Measured in Bushels (U.S. States)



A cut-off point of a labor intensity of 0.3 is applied. Only those grains are included which in the agricultural census are measured in bushels (corn for grain, wheat for grain, oats for grain, barley for grain, sorghum for grain, soybeans for beans).

Figure 3. Factor Intensity and Productivity, Grains Measured in Bushels (U.S. States), in natural logarithms



A cut-off point of a labor intensity of 0.3 is applied.

Figure 4. Factor Intensity and Productivity, All Grains (U.S. States)