**Appendix A, Methodology and Estimation Strategy**

The Cumulative Sum (CUSUM) type index (including CUSUM and CUSQ etc.) is an important tool used to identify the stability of a statistical process. This is achieved by examining the changes in the variance of a univariate time series or the residual obtained from a series of linear regressions of the time-series variable on its regressors with general mixing assumptions (strong correlation between regressors and unobserved error terms) [19,26-30].

To see how the index works, we note that any unanticipated upwards or downward (parameter) shifts will result in a uni-directional drift of the cumulative sum of variances or residuals over time. Thus, the CUSUM procedures give an out-of-control signal, when the absolute values of the cumulative sum exceed a critical value, indicating that the variable’s values after a specific time point have been significantly different from their previously expected levels. When a significant out-of-control signal occurs, the starting point of when the process goes out-of-control is determined and the magnitudes of parameter shifts are also estimated. The CUSUM control schemes are designed to optimally detect out-of-control states, and can therefore outperform other statistical techniques in this type of analysis.

**Mathematical Explanation of the CUSUM Index Using a Prototype Model**

Following Brown, Durbin and Evans [26], a prototype model of structural change analysis generating a CUSUM index assumes there is a linear regression with regressors  such that

 (A1)

Where is a time-series vector, which can be explained by a group of factors,  and is the estimated coefficients of  from the OLS regression with the previous  observations.

If the total number of observations is, the recursive residual at time  can be estimated by using the  observations.

 (A2)

wherecontains factors that determine . Thus, the variance of the predicted residual at time can be written as: .

If we choose a random time point  in the process such that , the th standardized recursive residual from Equation (1) can be defined as:

 (A3),

and the mean of this series of standardized residuals is given by  (where  is the number of regressors).

If the coefficients in the whole sample period stay constant, we have  to be normally distributed, where  is independent of  (given ).We can accumulate these standardized residuals deflated by their standard error over the observation period (the CUSUM index) or estimate the relative sum of squared recursive standardized residuals (the CUSQ index). The comparison between the estimated CUSUM/CUSQ index and their corresponding pre-determined criteria can be used to tell the potential structural break in time-series  under various conditions.

The CUSUM Index can be defined as:

 (A4)

Where wr is the th standardized recursive residual and  is the standard error of the series.

Under the initial assumption, the mean of CUSUM index is zero and variance is approximately equal to the number of accumulated residuals. The pre-determined boundary criteria are set at [] and  where *a* is the statistical significance level..If the CUSUM index (estimated with Equation (A4)) at time lies outside these boundaries it suggests that there is a possible breaking point in time-series () at time .

The CUSMSQ index can be defined as:

 (A5)

Where the expected value of St,, approximates to , which satisfies the Chi-square distribution. The pre-determined criteria is set between and , where is the statistical significance level . If there is a structural break, St is likely to lie outside of the interval between and .

**Adjusted CUSUM Squares Test under General Mixing Conditions**

The prototype model of the CUSUM/CUSQ test can be used to implement the structural change analysis of time-series variables. However, it has been widely criticised for not being able to address the case when there is a strong correlation between regressors and unobserved error terms. To address this issue, Tang and MacNeill [28] first adjusted the CUSUM/CUSQ test to account for serial correlation among error terms (in Equation (1)). Later, Bai and Perron [22] further extended the CUSUM test to cover the issue of identifying multiple structural changes for a given period. More recently, Deng and Perron [19] improved the CUSUM/CUSQ test by developing an adjusted CUSQ test that is independent of backward coefficient estimation () especially when there are limited observations. The advantages of the new method are that (1) it can be used to identify the structural change even when a weakly stationary situation exists (where ); (2) it can be used to identify multiple structural breaks at the same time.

To see how the adjusted CUSQ index works, we follow Deng and Perron ([19]; p. 2) by assuming that (1) for , 

with  a non-random positive definite matrix; (2) If  (or Equation (1) is weakly stationary), then (a)  forms a strongly mixing sequence with size -4k/(k-2) for space k>2, (b)  and  for some , (c) Let  for each  of length 1,  for some function v as  (with  the inner product). (3) Assumption (2) holds with  replaced by or , where .

Based on these three assumptions, Deng and Perron [19] proposed and proved an adjusted CUSQ test, which is invariant to non-Normal errors, serial correlation and conditional heteroskedasticity, as below:

 (A6)

with and . denotes either the OLS (), or the recursive residuals (), and  is a weight function and  some bandwidth which can be elected using one of the many alternative ways that has been proposed by Andrews [17]. Also the limit distribution is  , and the critical values at the 1 per cent, 5 per cent and 10 per cent are 1.63, 1.36 and 1.22, respectively.

Equations (A5) and (A6) are comparable in concept, since the denominator in both equations are the sum of the squared recursive residuals to time  and the numerator the sum of the squared recursive residuals over the entire observation period. However, the denominator and the numerator in Equation (A6) are adjusted for serial correlation, heteroskedasticity and a limited number of observations by using various weighting functions. Since the adjusted CUSQ index(defined in Equation (A6)) can be estimated independently of regression technique (either are cursive or OLS regression would be fine in this case) and immune to endogeneity between regressors and error terms, it provides a useful tool for identifying the structural break of time series. In addition, the adjusted CUSQ index can be used to identify more than one structural break at the same time since it adjusts the scale in denominator, which makes it superior to other methods in the structural analysis.

**Estimation Strategy for Identifying Agriculture TFP Structural Change**

To examine structural changes in broadacre TFP in Australia and its determinants, we adopted a multi-step testing procedure. The first step was to regress the logarithm of estimated broadacre TFP on a time trend (i.e. year), and use the residuals obtained from the regression to calculate the adjusted CUSQ index. The regression function can be written as:

 (A8)

Where  is the logarithm of broadacre TFP at time , and  is a time variable.

The second step is to incorporate climate and a knowledge stock, representing technologies available to farmers, into the OLS regression and re-calculate the adjusted CUSQ index. The regression function can be written as:

 (A9)

where is the climate variable (approximated by using the water stress index for the cropping industry), and (approximated as the weighted sum of public investment in agricultural R&D where can be 35 or 16 representing different time lags over hwihc new technologies become available and used by farmers) is the knowledge stock. Since there is little known about the length of lags between research and adoption by farmers and about the shape of the research lag profile, we have followed Mullen [4] in testing two alternative knowledge stock variables. Comparing results obtained from the first and second steps shows the extent to which climate change and agriculture R&D investments have affected the stability of agriculture productivity.

The third step is to incorporate education and terms of trade variables into the structural break analysis so that the impact of these factors on the stability of productivity growth could be examined. The regression function can be written as:



where is the education index, which is defined as the ratio of school attendance to the total population aged 4-19; and  is the logarithm of terms of trade for Australian broadacre agriculture.

Finally, we conduct a similar testing regime using industry-level and state-level data (as a sensitivity test) to assess the robustness of our results obtained from the aggregate data. However, due to data availability, the industry-level and state-level studies are only assessed using the first step (Equation (8)).

**Table A1:** OLS Estimation Results for the CUSQ procedure.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Base model | | R&D 35 lag | | | R&D 16 lag | | |
|  | time | time, lnsis | time, lnsis, r&d32 | time, lnsis, r&d32, educ | time, lnsis, r&d32, educ, lntot | time, lnsis, r&d16 | time, lnsis, r&d16, educ | time, lnsis, r&d16, educ, lntot |
| Year | 0.020\*\*\* | 0.019\*\*\* | 0.008\*\*\* | 0.007\*\* | 0.004 | 0.012\*\*\* | 0.010\*\*\* | 0.009\*\*\* |
|  | (0.001) | (0.001) | (0.002) | (0.003) | (0.003) | (0.002) | (0.002) | (0.003) |
| lnWt | - | 0.277\*\*\* | 0.279\*\*\* | 0.276\*\*\* | 0.271\*\*\* | 0.287\*\*\* | 0.284\*\*\* | 0.262\*\*\* |
|  | - | (0.052) | (0.043) | (0.044) | (0.042) | (0.043) | (0.043) | (0.042) |
| Kt35 | - | - | 0.196\*\*\* | 0.200\*\*\* | 0.144\*\*\* | - | - | - |
|  | - | - | (0.041) | (0.041) | (0.047) | - | - | - |
| Kt16 | - | - | - | - | - | 0.185\*\*\* | 0.189\*\*\* | 0.122\*\*\* |
|  | - | - | - | - | - | (0.038) | (0.038) | (0.044) |
| educ | - | - | - | 0.010 | -0.234\*\* | - | 0.010 | 0.009 |
|  | - | - | - | (0.011) | (0.108) | - | (0.011) | (0.010) |
| lntot | - | - | - | - | -3.208\*\*\* | - | - | -0.213\*\* |
|  | - | - | - | - | (1.007) | - | - | (0.106) |
| constant | 0.015 | -1.243\*\*\* | -3.320\*\*\* | -4.111\*\*\* | -3.208\*\*\* | -3.222\*\*\* | -4.035\*\*\* | -1.935 |
|  | (0.027) | (0.237) | (0.473) | (0.950) | (1.007) | (0.450) | (0.936) | (1.254) |
| R-square | 0.906 | 0.938 | 0.957 | 0.957 | 0.960 | 0.957 | 0.957 | 0.962 |
| Num. of Obs. | 55 | 55 | 55 | 55 | 55 | 55 | 55 | 55 |

**Note:** Standard errors in parentheses, and “\*\*\*”, “\*\*” and “\*” represent the significance at 1 per cent, 5 per cent and 10 per cent level respectively.

Source: Authors’ own calculation.

**Appendix B, Data Collection and TFP Estimation**

The aggregate TFP used in this analysis is estimated with the farm-level data mainly coming from ABARES’ Australian Agricultural and Grazing Industries Survey (AAGIS) of broadacre farmers, which has been carried out annually since the financial year of 1953. The method used for the estimation is the two-period non-transitive Fisher Indexing approach.

Fisher’s ideal index: Index numbers are usually computed over time, but they can also be evaluated between farms and between time points. In the following let the subscripts *s* and *t* represent two farms at two, possibly different, time points.

The Fisher ideal quantity index is defined in terms of the Laspeyres and Paasche quantity indexes:

and (B1)

respectively, where 1 ≤ *i* ≤ *N,* represents the input (output) used (produced) on the farm, *N* is the total number of inputs (outputs), *pis* is the price for input (output) *i* for *s* and *qis* is the quantity used (produced) of *i*.

Fisher's ideal index is the geometric mean of the Laspeyres and Paasche indexes:

 (B2)

When applying the formulae above, a farm at a given time point is chosen as the base and given an index value of 1. The index values of all other farms at all other time points are then relative to this base farm. If an index number is transitive any two farms at any two time points can be compared to each other by simply dividing their respective index numbers. In other words, if the index is transitive, the estimates from a regression analysis will not depend on which farm is chosen as the base. This is critical when working with farm level data.

To ensure that Fisher's index is transitive the following transformation is applied:

 (B3)

where 1 ≤ *r* ≤ *n* and *n* is the number of sample points, including repeated observations of the same farm at different time points.

Total factor productivity is a ratio of an output quantity index relative to an input quantity index. Each index represents the change in combined inputs or combined outputs relative to a chosen farm or time. The input or output quantities are aggregated using corresponding market prices. The groups of variables making up the inputs and outputs in the present report are described below

**Data Collection for Inputs and Outputs**

### Inputs

Inputs consist of 28 items that can be split into five major groups: land, capital, livestock purchases, labour, material and services.

**Land** The value variable for land is the opportunity cost of investing funds in this capital item. This is calculated as the average capital value (that is, the average of the opening and closing values) multiplied by a real interest rate. The quantity variable used for land is the area operated.

**Capital** is divided into plant and machinery, structures and livestock. The value variable for livestock is the opportunity cost of investing funds in this item. This is calculated as the average capital value (that is, the average of the opening and closing values) multiplied by a real interest rate. The value variables for structures, plant and machinery are the opportunity costs plus depreciation. For beef cattle, sheep and other livestock, the quantity variable is the average of opening and closing numbers. For buildings and plant capital, it is the average value of capital stock deflated by the respective prices paid indexes for each.

Livestock purchases, as a subgroup of capital, are split into beef, sheep and other livestock purchases. Their value equals purchases plus the value of natural changes (births and deaths) and transfers out provided together these are negative. The quantity variables for sheep and beef are derived from the respective value variables and respective prices received indexes for sheep meats and slaughtered beef. For the relatively small category of other livestock, the quantity variable is derived from the value of purchases and prices received index for livestock products.

**Labour** consists of four items - owner operator and family labour, hired labour, shearing costs, and stores and rations. The value of the owner operator and family labour input is imputed using weeks worked (collected as part of the survey) and an award wage. The value of hired labour is wages paid, and the values of shearing and stores and rations are expenditure on these items. The quantity variables for owner operator and family labour and hired labour are weeks worked. Expenditure on shearing was deflated by a shearing prices paid index is the quantity variable for shearing.

There are seven items in the **materials and services** group – fertiliser, fuel, crop chemicals, livestock materials, seed, fodder and other materials – and there are eight items in the services group – rates and taxes, administrative costs, repairs and maintenance, vet expenses, motor vehicle expenses, insurance, contracts and other services. The value for each item is expenditure. The quantity variables are derived by deflating the expenditure on each by the appropriate prices paid index.

### Outputs

Output consists of twelve items, which can be divided into four major groups: crops, livestock sales, wool and other farm income.

**Crops** are split into wheat, barley, oats, grain sorghum, oilseeds and other crops. The value variable for wheat is the quantity harvested multiplied by the Australian Wheat Board’s average net return for that year’s pool. For other grains and other crops, the value variable is net receipts in that year. The quantity variable for each of the grains is the quantity harvested. For the other crops, it is receipts deflated by the prices received index for crops.

For **livestock sales** of beef, sheep and lambs, the value is sales plus the value of natural changes (births and deaths) and transfers in provided together these are positive. For the minor category of other livestock, the value variable is sales. The quantity variables for beef, sheep and lambs are derived from the respective value variables and the prices received indexes for slaughtered beef, sheep and lamb meats. For the category of other livestock, the quantity variable is derived from the value of sales and a prices received index for livestock products.

For **wool** the value variable is net receipts. The quantity variable is the amount of wool shorn in kilograms.

And finally for **other farm income** the value variable is receipts and the quantity variable is receipts deflated by the sector prices received index.

**Definition of Other Variables**

In addition to the estimated TFP at the aggregate level, we also use other four variables as independent variables in the OLS regressions to examine their potential impact on the stability of productivity growth over time.

 is the logarithm of an index of crop water stress for broadacre industry, which is used to account for the impact of climate (both variability and climate change)Due to data availability, we construct this variable with different data sets over time. Specifically, for the period of 1953 to 1988, the index is obtained directly from APSRU, CSIRO. However, for the period of 1989 to 2004 (since there is no aggregate level data available), we have to use the weighted average of the water stress index at the farm-level (with DSE as the weight) to estimate its growth rate at the aggregate level. Thereafter, we apply the aggregate level growth rate of the water stress index to its previous trend and update its absolute level. For the period of 2004 to 2007 (even the individual level data is not available), we take the weighted average of the total rainfall at the farm-level (with DSE as the weight) to estimate the growth rate at the aggregate level. Thereafter, we apply the aggregate level growth rate of the total rainfall to the previous trend of the water stress index and update its absolute value.

There is little by way of theory or empirical evidence to guide the construction of knowledge stock variables representing the technology available to farmers except that most empirical work suggests that these lags are long [31]. Mullen [3] used a trapezoidal 35 year lag research profile,  and a 16 year inverted V profile, , to proxy the impact of public investment in research and development. Mullen’s database on R&D investment has been updated to 2007 from ABS sources.

 is defined as the percentage of enrolled students in schools in the total population aged between 4 and 19. The index was initially used by Hastings, and then updated by Mullen and Cox [3]. We used the same methodology. Enrolment is defined as "school attendance" or "the number of school students at the national level" obtained from ABS 4102.0 (various years: 2005, 2009) for the period of 1994-95 to present. The population data came from ABS 3201.0 Population by Age and Sex, Australian States and Territories (TABLE 9.1 Estimated Resident Population by single year of age, Australia). To smooth the trend, we take the 5-year moving average of the data series.

is defined as the logarithm of terms of trade for broadacre agriculture derived from ABARES data.

Finally, some descriptive statistics of the four variables are shown in Figures B1-B4.



**Figure B1:** Comparison of “crop water stress index” and “pasture grass growth index”: 1952-53 to 2006-07



**Figure B2:** Trends in public real agriculture R&D investment with 35-year and 16-year lags and public extension index: 1952-53 to 2006-07



**Figure B3:** Trends in education index in Australia: 1952-53 to 2006-07



**Figure B4:** Trends in broadacre terms of trade: 1953 - 2007