Understanding the Drivers of Productivity through Regression Analysis

Introduction

Productivity and its spatial variability are central to the Rural Economy Growth Review, conducted in Autumn 2011. Complementary analysis has been carried out to establish a statistical understanding of the factors that determine productivity and in particular the role played by spatial considerations such as rurality and connectivity. This investigation builds upon an econometric study by Agarwal, Rahman and Errington to produce a group of empirical models. The first section of the report offers a non-technical summary of the key findings, whilst the latter section documents the outputs of the different models, their limitations, and a more detailed interpretation.

Non-technical summary of findings

This report seeks to identify the set of factors that best explain productivity (expressed as Gross Value Added per Work Force Job {GVA per WFJ}). Traditionally, productivity is usually thought to be driven by capital (including finance, skills, infrastructure and ICT), labour and innovation. New Economic Geography and Agglomeration theory have also suggested that concentrations of economic activity generate economic benefits for the firms located within them[1]. The results of this study show finance and innovation to be major factors in determining productivity and these are supported by several factors which relate to infrastructure. It is interesting that skills and ICT have not proved to be prominent; however it is likely that these factors have been captured indirectly within other drivers. Rurality is not usually a consideration in assessments of productivity. This investigation found that productivity was higher where there were larger village and 'Large Market Town' populations, both rural settlement types. However there was no evidence that simply being in a rural rather than urban LA leads to differences in productivity.

It is important to recognise that the analytical techniques used here (regression) can help us to understand the relationships between productivity (expressed as Gross Value Added per Work Force Job {GVA per WFJ}) and its drivers, but will not be able to explain them exactly with absolute certainty. For example, our productivity proxy is only available at Local Authority district level and therefore reflects an average across quite a wide range of places (and is also smoothed across a number of years). This method and limitations of the data mean that only some, not all of the variation in productivity can be explained. The relationships that can be explored using regression analysis do not prove causation, however part of the variable selection process is ensuring that the drivers implied by the model make intuitive sense.

[1] For a summary of this literature, see http://www.bis.gov.uk/assets/biscore/economics-and-statistics/docs/u/10-1226-understanding-local-growth

Four key variables were identified as being robust drivers of productivity that we can be most sure of have a strong influence:

Table 1: The four 'key' variables identified as drivers of productivity

Variable number	Variable description	Drives productivity
1	Business start ups per 1,000 population	▲
2	Number of employees per business unit	▲
3	Capital investment per work force job	▲
4	Proportion of employees who are in public service	▼

Of these, Business start ups per 1,000 population and Number of employees per business unit appear to be the most reliable drivers and Proportion of employees who are in public service the weakest of the four. The most notable observation from a rural perspective is that none of these are spatial factors. This tells us that it is not primarily location that determines productivity, but characteristics of local businesses. This does not mean that productivity or business characteristics do not vary between rural and urban areas, but that the variability in productivity is *best* explained by the variables above rather than by specifically spatial variables.

A further four variables were identified as having a noticeable effect on productivity. Both the key and additional variables are shown in Table 2 below, ordered beginning with the most robust.

Table 2: The four 'key' variables and four additional variables identified as drivers of productivity

Variable number	Variable description	Beta (sensitivity)	Standardised Beta (impact)
1	Business start ups per 1,000 population	0.156	0.231
2	Number of employees per business unit	0.321	0.488
3	Capital investment per work force job	0.084	0.179
4	Proportion of employees who are in public service	-0.129	-0.134
5	Percentage of population living in Large Market Towns	0.014	0.185
6	Distance from City of London	-0.051	-0.265
7	Percentage of population living in villages	0.020	0.221
8	Number of centres of employment with >5000 jobs accessible within a 'reasonable' time by car	0.129	0.182

The Beta column shows the estimated percentage change in GVA per WFJ from a 1% increase in the variable. For example, if business start ups per 1,000 population increased by 1% (either over time or comparing one LA to another), we would expect to see a 0.156% increase in GVA per WFJ, all other things being equal. Similarly, if business start ups per 1,000 population increased by 10%, we would expect to see productivity increase by approximately 1.6%. Negative betas, as seen for two of the variables mean that we would expect GVA per WFJ to fall in response to an increase in either of these. The larger the beta, the more sensitive productivity is to a change in the variable. However this needs to be considered in conjunction with the magnitude of changes likely for each variable. The

Standardised Beta column helps to demonstrate this. It shows the sensitivity of each variable relative to its variability (standard deviation). This allows a fairer comparison between variables that tend to be spread over different ranges and can be thought of as the impact of each variable on productivity. For example, productivity is not very sensitive to changes in the percentages of population in Large Market Towns and villages (variables 5 and 7), shown by small Betas (0.014 and 0.020), but is more sensitive to changes in the proportion of employees who are in public service (Beta of -0.129). However the proportion of employees who are in public service has a very low variability, tending to be between 20% and 30% in most LAs with very few extreme values. In contrast, the Large Market Town and village variables have a much greater spread of values and many zeros. As a result, the Large Market Town and village variables explain more of the variability in productivity than public service, shown by their higher Standardised Betas of 0.185 and 0.221, compared to (-)0.134. Using this measure, Number of employees per business unit (0.488) and Distance from City of London (-0.265) have the biggest impact on GVA per WFJ, however all of the variables here are making a noticeable contribution. It is important to note that whilst the four additional variables (5-8) have a similar impact to the key variables (1-4), we are far less certain of the accuracy of this impact. The additional variables can be replaced with other variables which seem to explain similar effects on productivity, whereas the inclusion of the key variables is critical to the overall explanatory power of the model.

Whilst the 'key variables' included no spatial elements, all of the additional variables are location related. The Distance from City of London variable suggests that productivity will fall as distance from the capital increases. GVA has been adjusted for regional prices, which lessens the impact of higher values in London and surrounding regions where prices are also highest. Clearly, the distance from an LA to London cannot be changed, however this is likely to be a function of connectivity and economies of scale, issues which could be addressed. The model shows that productivity increases in areas where a higher proportion of the population live in 'large market towns' or villages, although the effect is small. We do not know whether GVA is being generated in these settlements or whether they provide the living environments that attract and retain highly productive workforces. Variable number 8 is intended to capture the influence of accessibility on productivity, with areas open to more large centres of employment delivering higher productivity. Access to a wider selection of jobs should lead to greater flexibility and efficiency, driving productivity upwards.

The four key variables are able to explain just short of 40% of the variability in productivity which improves to 47% following the introduction of the four additional variables. This highlights the limitations of the explanatory powers of this method; however these levels are reasonable considering the complexity of the system being modelled.

Technical Report

Variables and data

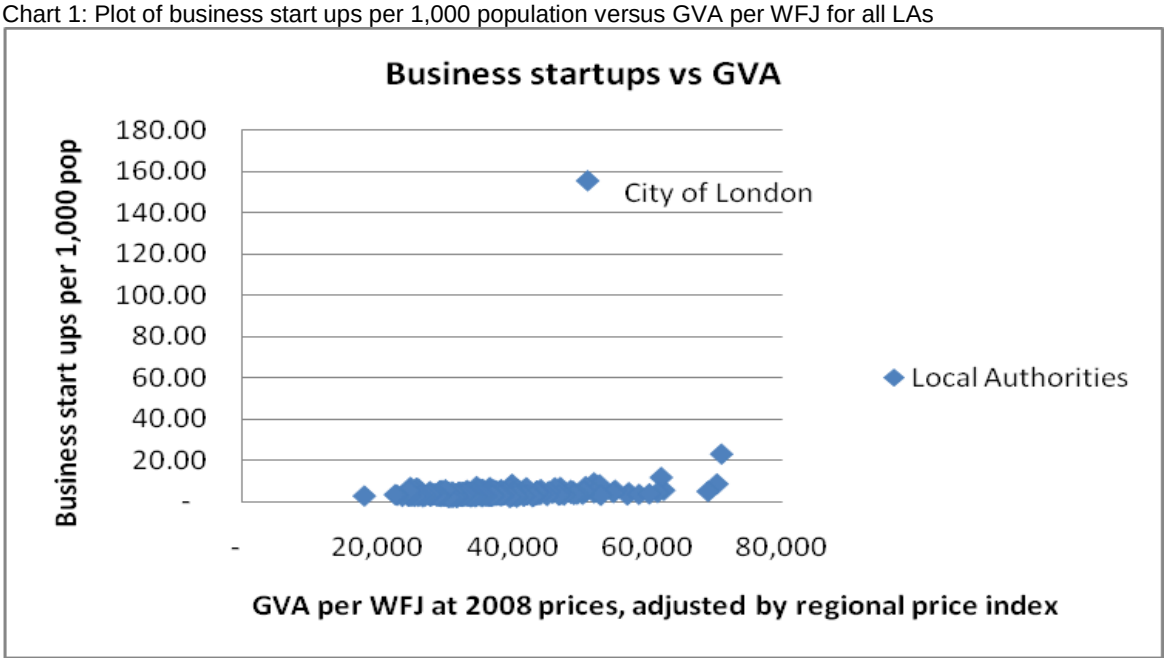
The dependent variable is gross value added (GVA) per work force job (WFJ) at local authority (LA) level. All LAs (post 2009 revisions) in England were used in the initial analysis except for The Isles of Scilly which lack data for many vital measures. A straight average of GVA per WFJ was taken over 2001-08 to smooth significant inter-year fluctuations which are known to exist in the GVA estimates at LA level. The GVA values from each year were adjusted to 2008 prices. Finally, each LA figure was adjusted to account for regional price differences using a combination of two regional price indices devised by ONS (no price index at a lower geographic level was available).

The explanatory variables were chosen from over 130 proposed variables covering a broad range of topics including economic activity, connectivity, enterprise structure, settlement structure, skills and deprivation. The data has been taken from a number of sources and the most recent data available has tended to be used. This means that the explanatory variables do not correspond precisely to the measurement period for the dependent variable and in some cases fall outside it completely. Concerns remain over the accuracy of the GVA data, which ONS are reluctant to allow use of at individual LA level. This is a weakness of this analysis and will reduce the ability of the models to explain all of the variability in GVA.

Linear regression analysis was performed in PASW Statistics 18 (IBM SPSS). All variables, both dependent and explanatory were logged, both to manage extreme values and allow the results to be interpreted as elasticities. This required small values to be added to variables which has some zero values (as 0 cannot be logged).

Special attention was paid to the City of London LA, which has some unique qualities leading to unusual data points. The City of London has the smallest population of any LA and is highly business focused. It has a high GVA per WFJ, being ranked 14th out of 325 LAs, but is not at the top as might be expected. The biggest impact of the City of London was the massive outlier that it represents for the 'business start ups per 1,000 population' variable. The City of London's value of 155.65 is 6.7 times the size of the next largest observation (23.18) and almost 40 times the average of other observations (3.98). This is represented visually by Chart 1 below.

Chart 1: Plot of business start ups per 1,000 population versus GVA per WFJ for all LAs



The 'business start ups per 1,000 population' variable is very important for explaining productivity. Once the outlying observation for the City of London is constrained to below 105, 'business start ups per 1,000 population' becomes the dominant variable suggested by a step-wise approach, displacing the 'distance to London' variable which measures the distance from the population centroid of each LA to the population centroid of the City of London. Due to the excessive influence of the City of London on the regression process and it not being a good representative of other LAs, it was decided to exclude the City of London from the final analysis.

Method of Model Identification

A stepwise approach was run on the entire set of LAs (minus the City of London) to get a quick view on the most influential variables. The set of 324 LAs was randomly split into two thirds (215 LAs) and one third (109 LAs). The stepwise method was used on the two thirds and a suitable looking model selected from the results, striking a balance between a high R square (goodness of fit) and low p-values for each variable. It should be noted that the stepwise method used here biases p-values, so only exceptionally low p-values (<1%) were considered appropriate. The selected model was applied against the remaining one third of LAs. This process was repeated five times (i.e. five different random splits were taken and a new model selected each time). From the results, four 'key variables' were identified. These were variables that had appeared in at least three of the five trial models and also ranked highly in the step-wise models suggested for all 324 LAs. Table 3 below shows the four key variables.

Table 3: The four 'key' variables identified as drivers of productivity

Variable name	Variable description
ln_num_bus_startups_per_1000pop	Business start ups per 1,000 population
ln_num_employees_per_localunit	Number of employees per business unit
ln_num_CapInvest_per_WFJ	Capital investment per work force job
ln_pc_public_sector_pop	Proportion of employees who are in public service

Four additional variables were selected based on a combination of the results of the stepwise analysis described above, manually entering combinations of possible selections and whether the result was intuitive. The four additional variables are described in Table 4 below.

Table 4: The four additional variables identified as drivers of productivity

Variable name	Variable description
ln_pc_LMT_pop	Percentage of population living in Large Market Towns
ln_num_dist_London	Distance from City of London
ln_pc_Village_pop	Percentage of population living in villages
ln_num_emp5000_reasonable_newcar	Number of centres of employment with >5000 jobs accessible within a 'reasonable' time by car

Regression Results

1) The Basic Model

This model consists of only the four key variables explained above, plus a constant term.

PASW18 Outputs for the Basic Model:

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.627[a]	.393	.386	.15633

a. Predictors: (Constant), ln_pc_public_sector_pop, ln_num_employees_per_localunit, ln_num_CapInvest_per_WFJ, ln_num_bus_startups_per_1000pop

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients		
		B	Std. Error	Beta	t	Sig.
1	(Constant)	8.624	.173		49.908	.000
	ln_num_bus_startups_per_1000pop	.272	.034	.404	8.066	.000
	ln_num_employees_per_localunit	.244	.031	.370	7.933	.000
	ln_num_CapInvest_per_WFJ	.090	.022	.192	4.089	.000
	ln_pc_public_sector_pop	-.172	.045	-.178	-3.838	.000

a. Dependent Variable: ln_num_GVA_per_WFJ_reg_prices

The results for the Basic Model show:

1. An R Square of 0.393 which means that this model is explaining 39.3% of the variability in the GVA per WFJ. This is acceptable, but highlights that this method of analysis can help us to understand the relationships between productivity and its drivers, but will not be able to explain them exactly.
2. Each of the four variables have a significance of 0.000 which means that they are all highly significant. We can be very confident that they each impact on productivity.
3. The betas in the "B" column of the results show the estimated percentage change in (unlogged) GVA per WFJ from a 1% increase in the (unlogged) variable.
4. The negative beta for ln_pc_public_sector_pop shows that we would expect GVA per WFJ to fall given an increase in public sector employment.
5. The Standardized Coefficients give an indication of impact on productivity, giving more weight to variables that tend to vary more than others. Business start-ups per 1,000 population has the largest Standardized Beta (0.404), closely followed by Employees per local unit (0.370).
6. Standardised residuals were plotted in a histogram which fits the Normal Distribution well. This is an indicator of reliable results. The residuals are the difference between the actual observed values of the dependent variable (GVA per WFJ) and those predicted by the model.

7. The values of GVA per WFJ predicted by the model were plotted against their residual values. There is a no clear correlation, which is a further indicator of reliable results as it demonstrates homogeneity of variances.

2) The Complex Model

This model uses the four key variables and introduces terms to explain the relationships between them.

Further testing was carried out to understand the suitability of these variables and their relationship to the dependent variable. The correlation table below shows low levels of correlation (maximum 0.28) between the four key variables. This should mean that these four variables are mostly explaining different rather than overlapping underlying processes. The correlations between each key variable and the dependent variable are in general larger, but are all <0.5 which suggests that the model is unlikely to be able to explain a large proportion of the variability in GVA per WFJ.

Table 5: Correlations between the four unlogged 'key' variables and the dependent variable. Correlations between the key variables are fairly low, which is helpful.

	GVA per WFJ	Business start ups per 1,000 population	Number of employees per business unit	Capital investment per work force job	Proportion of employees who are in public service
GVA per WFJ	1.00	0.46	0.24	0.41	-0.29
Business start ups per 1,000 population	0.46	1.00	-0.16	0.28	-0.27
Number of employees per business unit	0.24	-0.16	1.00	0.16	0.12
Capital investment per work force job	0.41	0.28	0.16	1.00	-0.16
Proportion of employees who are in public service	-0.29	-0.27	0.12	-0.16	1.00

Quadratic terms (produced by squaring each logged value of a variable) were tested for each of the four key variables. Each of the quadratic terms proved to be significant at first. Six interaction variables were also created as the products of (logged) key variables. Once the interaction terms were introduced to the regression, the quadratic variables ceased to be significant so were removed. A backward step-wise procedure was used to identify which interaction variables were most significant. The three interaction variables selected using this process are described in the Table 6 below.

Table 6: The three interaction terms selected for the complex model

Variable name	Variable description
int_ln_num_employees_per_localunit_ln_pc_public_sector_pop	Interaction between Number of employees per business unit and Proportion of employees who are in public service
int_ln_num_employees_per_localunit_ln_num_CapInvest_per_WFJ	Interaction between Number of employees per business unit and Capital investment per work force job
int_ln_num_CapInvest_per_WFJ_ln_num_bus_startups_per_1000pop	Interaction between Capital investment per work force job and Number of business start ups per 1,000 population

PASW18 Outputs for the Complex Model

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.708[a]	.502	.491	.14234

a. Predictors: (Constant), int_ln_num_CapInvest_per_WFJ_ln_num_bus_startups_per_1000pop, int_ln_num_employees_per_localunit_ln_num_CapInvest_per_WFJ, ln_pc_public_sector_pop, ln_num_CapInvest_per_WFJ, ln_num_employees_per_localunit, ln_num_bus_startups_per_1000pop, int_ln_num_employees_per_localunit_ln_pc_public_sector_pop

b. Dependent Variable: ln_num_GVA_per_WFJ_reg_prices

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients		
		B	Std. Error	Beta	t	Sig.
1	(Constant)	8.678	.159		54.520	.000
	ln_num_bus_startups_per_1000pop	.170	.034	.252	5.057	.000
	ln_num_employees_per_localunit	.138	.031	.210	4.403	.000
	ln_num_CapInvest_per_WFJ	.067	.021	.143	3.250	.001
	ln_pc_public_sector_pop	-.105	.043	-.109	-2.443	.015
	int_ln_num_employees_per_localunit_ln_pc_public_sector_pop	-.059	.019	-.168	-3.121	.002
	int_ln_num_employees_per_localunit_ln_num_CapInvest_per_WFJ	.011	.004	.156	2.854	.005
	int_ln_num_CapInvest_per_WFJ_ln_num_bus_startups_per_1000pop	.020	.004	.262	5.619	.000

a. Dependent Variable: ln_num_GVA_per_WFJ_reg_prices

The results for the Complex Model show:

1. An R Square of 0.502 which means that this model is explaining 50.2% of the variability in the GVA per WFJ.
2. All seven variables have small significance p-values which mean that they are all highly significant. We can be very confident that they each impact on productivity. The largest p-value is 1.5% for the public sector employment variable. This is still small but as all four key variables appear in at least one of the three interaction terms, it is only the p-values of the interaction terms which are relevant.
3. The betas in the "B" column of the results show the estimated percentage change in (unlogged) GVA per WFJ from a 1% increase in the (unlogged) variable. For an interaction term, the beta is the change expected from a 1% increase in both underlying variables in addition to the individual betas of these variables. For example, if Number of business start ups per 1,000 population and Capital investment per WFJ each increased by 1%, we would expect GVA per WFJ to increase by (0.170%+0.067%+0.020%=) 0.257%. This suggests that capital investment and business start ups are reinforcing.
4. The Standardized Coefficients give an indication of impact on productivity, giving more weight to variables that tend to vary more than others. Similar to the Unstandardized Beta, this is complicated by the interaction terms. Business start-ups per 1,000 population has the largest Standardized Beta of the key variables (0.252) and also shares a large interaction term with the capital investment variable (0.262). The employees per business unit variable also has a large impact with a measure of 0.210.
5. Standardised residuals were plotted in a histogram which fits the Normal Distribution well. This is an indicator of reliable results. The residuals are the difference between the actual observed values of the dependent variable (GVA per WFJ) and those predicted by the model.
6. The values of GVA per WFJ predicted by the model were plotted against their residual values. There is a no clear correlation, which is a further indicator of reliable results as it demonstrates homogeneity of variances.

3) Extended model

This model uses the four key variables and four additional variables. Interaction variables have not been used for parsimony.

The four additional variables were identified by the stepwise analysis of the full 324 LA file, but they did not show as strong variables during further testing.

The full correlation table (Table 7 below) shows a maximum correlation of (-)0.59 between Percentage of population living in villages and Number of employees per business unit. Whilst this is not high enough to destabilise the model, it does make the impacts of individual variables difficult to separate, meaning that all the results in this report should be treated as indications of the true picture, rather than statements of fact.

Table 7: Correlations between the four unlogged 'key' variables, the four additional variables and the dependent variable. Some correlations involving additional variable are >0.5, which makes the regression results more difficult to interpret.

	GVA per WFJ	Business start ups per 1,000 population	Number of employees per business unit	Capital investment per work force job	Proportion of employees who are in public service	Percentage of population living in Large Market Towns	Distance from City of London	Percentage of population living in villages	Number of centres of employment with >5000 jobs accessible within a 'reasonable' time by car
GVA per WFJ	1.00	0.46	0.24	0.41	-0.29	-0.10	-0.33	-0.17	0.41
Business start ups per 1,000 population	0.46	1.00	-0.16	0.28	-0.27	-0.04	-0.49	-0.04	0.27
Number of employees per business unit	0.24	-0.16	1.00	0.16	0.12	-0.47	0.16	-0.59	0.48
Capital investment per work force job	0.41	0.28	0.16	1.00	-0.16	-0.11	-0.19	-0.09	0.20
Proportion of employees who are in public service	-0.29	-0.27	0.12	-0.16	1.00	-0.09	0.38	-0.03	-0.05
Percentage of population living in Large Market Towns	-0.10	-0.04	-0.47	-0.11	-0.09	1.00	0.12	0.54	-0.47
Distance from City of London	-0.33	-0.49	0.16	-0.19	0.38	0.12	1.00	0.14	-0.25
Percentage of population living in villages	-0.17	-0.04	-0.59	-0.09	-0.03	0.54	0.14	1.00	-0.57
Number of centres of employment with >5000 jobs accessible within a 'reasonable' time by car	0.41	0.27	0.48	0.20	-0.05	-0.47	-0.25	-0.57	1.00

PASW18 Outputs for the Extended Model:

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.683[a]	.467	.453	.14751

a. Predictors: (Constant), ln_num_emp5000_reasonable_newcar, ln_pc_public_sector_pop, ln_num_CapInvest_per_WFJ, ln_pc_LMT_pop, ln_num_bus_startups_per_1000pop, ln_num_employees_per_localunit, ln_pc_Village_pop, ln_num_dist_London

b. Dependent Variable: ln_num_GVA_per_WFJ_reg_prices

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients		
		B	Std. Error	Beta	t	Sig.
1	(Constant)	9.207	.263		35.013	.000
	ln_num_bus_startups_per_1000pop	.156	.045	.231	3.503	.001
	ln_num_employees_per_localunit	.321	.040	.488	7.995	.000
	ln_num_CapInvest_per_WFJ	.084	.021	.179	4.026	.000
	ln_pc_public_sector_pop	-.129	.043	-.134	-3.007	.003
	ln_pc_LMT_pop	.014	.004	.185	3.393	.001
	ln_num_dist_London	-.051	.013	-.265	-3.829	.000
	ln_pc_Village_pop	.020	.006	.221	3.432	.001
	ln_num_emp5000_reasonable_newcar	.129	.036	.182	3.565	.000

a. Dependent Variable: ln_num_GVA_per_WFJ_reg_prices

The results for the Extended Model show:

1. An R Square of 0.467 which means that this model is explaining 46.7% of the variability in the GVA per WFJ. This is an improvement over the Basic Model (0.393), but lower than the Complex Model (0.502).
2. All eight variables have extremely small significance p-values which mean that they are all highly significant. We can be very confident that they each impact on productivity.
3. The betas in the "B" column of the results show the estimated percentage change in (unlogged) GVA per WFJ from a 1% increase in the (unlogged) variable. Comparing against the Basic Model, three of the key variables have smaller betas, however the beta for Employees per business unit has increased from 0.244 to 0.321.
4. The Standardized Coefficients give an indication of impact on productivity, giving more weight to variables that tend to vary more than others. Employees per local unit has the largest Standardized Beta (0.488). The distance from London variable has the next largest impact with a measure of -0.265.
5. Standardised residuals were plotted in a histogram which fits the Normal Distribution well. This is an indicator of reliable results. The residuals are the difference between the actual observed values of the dependent variable (GVA per WFJ) and those predicted by the model.
6. The values of GVA per WFJ predicted by the model were plotted against their residual values. There is a no clear correlation, which is a further indicator of reliable results as it demonstrates homogeneity of variances.

4) Complex Extended Model

This model uses the four key variables, the four additional variables and the three interaction terms. It is therefore a combination of the Complex Model and the Extended Model.

PASW18 Outputs for the Complex Extended Model:

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.745[a]	.555	.539	.13535

a. Predictors: (Constant), int_ln_num_CapInvest_per_WFJ_ln_num_bus_startups_per_1000pop, int_ln_num_employees_per_localunit_ln_num_CapInvest_per_WFJ, ln_pc_LMT_pop, ln_pc_public_sector_pop, ln_num_CapInvest_per_WFJ, ln_num_emp5000_reasonable_newcar, ln_num_dist_London, int_ln_num_employees_per_localunit_ln_pc_public_sector_pop, ln_pc_Village_pop, ln_num_employees_per_localunit, ln_num_bus_startups_per_1000pop

b. Dependent Variable: ln_num_GVA_per_WFJ_reg_prices

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	9.126	.243		37.579	.000
	ln_num_bus_startups_per_1000pop	.076	.042	.113	1.790	.074
	ln_num_employees_per_localunit	.198	.041	.301	4.881	.000
	ln_num_CapInvest_per_WFJ	.063	.020	.133	3.175	.002
	ln_pc_public_sector_pop	-.081	.041	-.084	-1.950	.052
	ln_num_dist_London	-.040	.012	-.209	-3.265	.001
	ln_pc_LMT_pop	.011	.004	.145	2.879	.004
	ln_pc_Village_pop	.017	.005	.186	3.122	.002
	ln_num_emp5000_reasonable_newcar	.132	.033	.186	3.963	.000
	int_ln_num_employees_per_localunit_ln_pc_public_sector_pop	-.048	.018	-.137	-2.643	.009
	int_ln_num_employees_per_localunit_ln_num_CapInvest_per_WFJ	.011	.004	.153	2.931	.004
	int_ln_num_CapInvest_per_WFJ_ln_num_bus_startups_per_1000pop	.019	.003	.250	5.607	.000

a. Dependent Variable: ln_num_GVA_per_WFJ_reg_prices

The results for the Complex Extended Model show:

1. An R Square of 0.555 which means that this model is explaining 55.5% of the variability in the GVA per WFJ. This is the highest R Square of the four models discussed in this report. This is to be expected as it has the most variables.
2. The three interaction terms and four additional variables have extremely small significance p-values which mean that they are all highly significant. We can be very confident that they each impact on productivity. The p-values of the key variables are not important in the presence of the interaction terms.
3. The betas in the "B" column of the results show the estimated percentage change in (unlogged) GVA per WFJ from a 1% increase in the (unlogged) variable. These are difficult to interpret for the key variables due to the interaction terms. This model does suggest that productivity is particularly sensitive to changes in ln_num_emp5000_reasonable_newcar, which related to the accessibility of large centres of employment.
4. The Standardized Coefficients give an indication of impact on productivity, giving more weight to variables that tend to vary more than others. Similar to the Unstandardized Beta, this is complicated by the interaction terms. Employees per local unit has the largest Standardized Beta (0.301), but also has a negative interaction term with the public sector employment variable. The distance from London variable has a large impact with a measure of -0.209.

5. Standardised residuals were plotted in a histogram which fits the Normal Distribution well. This is an indicator of reliable results. The residuals are the difference between the actual observed values of the dependent variable (GVA per WFJ) and those predicted by the model.
6. The values of GVA per WFJ predicted by the model were plotted against their residual values. There is a no clear correlation, which is a further indicator of reliable results as it demonstrates homogeneity of variances.

Other variables considered and rejected

Percentage of population living in major urban areas. This had a positive beta. It only proved highly significant when included alongside other variables not selected.

Number of people with no qualifications. This variables was highly significant, however it is not intuitive that a nominal variable should affect productivity per job. We would have expected percentage of working age population with no qualifications to be significant, however this was found not to be true.

Total number of employees (enterprise measure). This variable comes from the Inter Departmental Business Register (IDBR) where an enterprise is the smallest combination of legal units (generally based on VAT and/or PAYE records) which has a certain degree of autonomy within an Enterprise Group. This means that enterprise level data can represent several sites and is therefore not 100% specific to an individual LA. Similar to the Number of people with no qualifications, it is not intuitive that productivity per job should be affected by a nominal value.

Living Environment Deprivation Score. This variable is part of the Index of Multiple Deprivation (IMD). The indicators used in the latest update of this domain are; - Social and private housing in poor condition - Houses without central heating - Air quality - Road traffic accidents involving injury to pedestrians and cyclists. The higher the score, the more deprived, so it was not surprising that this variable had a negative beta. It showed up strongly in the stepwise analysis of the 324 LAs, but not when the file was split and the stepwise approach was used on randomly samples of two thirds of the file. If added to the Extended Model above, it does not prove significant. It could also be argued that a poor living environment may be experienced as a result of low productivity rather than being a driver.

Rural-Urban Dummy. Two simple binary measures of rurality were tried, with Significant Rural LAs classed as rural in one and as urban the other. When added to the Basic Model, these prove significant. However the Rural-Urban Dummy loses its significance to the additional variables once they are added.

Internet infrastructure. Broadband was an area that was of particular interest; however the data available was poor. The best variable at the required spatial level was 'Premises with Internet connection <2Mbit/s as % of Total Premises' from early 2011, well outside of the observation period for the dependent variable. This severely reduced its explanatory power as internet infrastructure improved significantly between 2008 and 2011 and reliance on broadband will have increased over the GVA measurement period of 2001-08. As a result, this variable was not significant in the regression models so was discarded.