Spring 4-1-2011

Target Industry Study: an Empirical Analysis of Intertemporal Trends in Regional Industrial Base Composition

David M. Smith
Lynchburg College

Target Industry Study: an Empirical Analysis of Intertemporal Trends in Regional Industrial Base Composition

Senior Research Project

Submitted in partial fulfillment of the graduation requirements
For the Economics major

Defended by

David M. Smith

April 2011

Reviewed by:

Dr. Dan Messerschmidt, Chair ______________________

Professor Mike Schnur ______________________

Dr. Eric Kyper ______________________

Target Industry Study: an Empirical Analysis of Intertemporal Trends in Regional Industrial Base Composition

David M. Smith

Senior Research Project

Submitted in partial fulfillment of the graduation requirements For the Economics major

School of Business and Economics

April 2011

I. Abstract

This paper follows existing literature on local industrial conditions and agglomeration economies in an adapted shift-share framework. The purpose of this study is to develop a cost-efficient empirical model that meets the needs of local government in analyzing changes in industrial composition. The data for this model is obtained from the Bureau of Labor Statistics' Quarterly Census of Employment and Wages (QCEW) with non-disclosed data provided by the Virginia Employment Commission which is sorted by North American Industrial Classification System (NAICS) 2-digit industrial sectors. The model developed in this paper identifies certain industrial sectors which exert significant influence on the local economic structure and serves as empirical support for adjustment of marketing campaigns and incentive packages.

II. Introduction

Economic development is a topic of interest for every territorial level. At each level, there are various interested parties who approach the same situation from different perspectives. Incumbent firms assess development initiatives by the posed threat of competition in both the sales of output and the cost of inputs. Residents are concerned with potential employment opportunities as well as the congestion externalities that accompany development. Incoming firms compare the incentive packages that are offered by competing territories and assess the local conditions of each possible site. Governing bodies maintain their focus on the effects that any development initiative will have on the tax base.

The model developed in this paper approaches the considerations for any proposed development from the perspective of the governing body. The changes in the tax base of a given territorial level can be used as a quick reference to measure the ability of the governing body to meet its budgetary obligations. This research will not follow the tax base changes but instead will incorporate the theoretical framework that is derived from fiscal limitations. Given that the financial resources of a governing body are limited, a profit-maximizing approach would dictate that any expenditure by the government would go towards the opportunity that offered the highest return. The incentive packages and marketing campaigns funded by the territorial government are investments that are intended to bring returns in the form of increased tax revenues on equipment and property as well as increased employment within the territory. The long-run indirect and induced effects of this investment are an anticipated increase in the tax base.

The dynamic nature of this model is intended to provide insight into the long-term effects that industrial sectors exert on the tax base. Through the examination of employment-weighted wages for each industrial sector, the natural advantages of the territory that favor certain industrial sectors will become evident. These natural advantages are intended to represent the available natural resources, the compatible infrastructure, the size and quality of the labor pool, complementary firms and industries, and the presence of competing firms and industries. All of these factors are assumed to have been considered by incumbent firms and are expected to be weighed by incoming firms as part of the decision to locate within the given territory. In the long-run, the industrial sectors that capitalize on these natural advantages will remain in the territory; those that are not appropriately suited will exit. In this model, the changes within sectoral employment are measured; firm entry and exit are not distinguished from firm expansion and contraction.

The criteria necessary to produce a model that is suitable for the selected perspective begins with minimizing the cost of the model, utilizing readily available data, and generating output which is easily interpretable. An ordinary least squares regression will be utilized in place of a more expensive input-output model. EViews 4.1 and EViews 7 are the software packages used to generate the statistical analysis of the variables developed in this paper.

The purpose of this study is to determine what, if any, is the observed historical contribution from each industrial sector within Campbell County to the overall economic structure of the locality. The null hypothesis for this model states that all employment-based earnings generated within the territorial unit exert an equal influence on the economic structure of the region. The alternative hypothesis claims that the employment-weighted wages in one industry will exert a different influence on the economic structure of the locality than the employment-weighted wages of the other sectors.

III. Literature Review

This investigation begins with the review of parallel research to determine the appropriate scope, the accepted modeling assumptions, and necessary considerations for data. Each included study contributes to this research through the development of Input-Output models, shift-share analyses, and various regression models which have all been employed to explain some aspect of the inter-industry relationships within a given location. This study is heavily dependent upon the shift-share analysis which has been utilized by several related studies as a preliminary calculation in their research.

The shift-share analysis is an accounting device that is used to analyze changes in a given attribute for the region relative to the aggregate. A shift-share, explained by D'Elia (2005), permits the observation of each industry within a region over time based on three components: the national share, the nation-wide industry mix, and the regional shift. The National Share (NS) represents the change in the given attribute that each regional sector would have experienced if it had exhibited the same pattern as the attribute nationwide. The Industry Mix (IM) represents the sectoral composition of the region and the specialization in the region of fast or slow changing industries at the aggregate level for the given attribute. The Regional Shift (RS) represents the local conditions that cause the regional sector's change to differ from the aggregate sector's change. "In a traditional shift-share model, a region with above average employment growth either has a favorable industry mix or it enjoys a competitive advantage over other regions" (Dinc & Haynes, 1997, p. 471). The traditional shift-share analysis compares only two points in time. The dynamic shift-share utilized by D'Elia (2005) analyzes the development of the components over time allowing the structural changes that are discounted in a two- period model to be examined. "Detecting these trends, similarities and differences as well as the underlying forces in the regional productive or employment structures is germane to a sound policy design towards enhancing regional economic performance" (D'Elia, 2005, p. 4).

The first obstacle introduced by the literature was the selection of the appropriate territorial level for analysis. Arauzo-Carod (2008) begins his analysis of firm entry decisions with the discussion of territorial level selection in which administrative units are compared to functional units. Administrative units are defined as a municipality or a county, whereas functional units are travel-to-work-areas (TTWA). Arauzo-Carod (2008) determines that most firm location decisions occur at the administrative territory level but also cautions that labor markets are not defined by municipality borders. The reliance on fixed-border units makes some level of sacrifice in accuracy to achieve a higher level of stability in the analyzed territorial unit. The identification of TTWAs is developed through the analysis of commuting patterns and Aruzo-Carod (2008) acknolwedged the inherent instability in this unit due to the changing nature of the individual's commuting patterns. Functional units are non-governmental and lack the authority to affect policy changes. Dinc & Haynes (1997) state that economic development policies, such as low interest loans, tax breaks, industrial recruiting, and investment in infrastructure, have been dominantly state and local issues. To this end, "sectoral structure and performance of a region is very important" (Dinc & Haynes, 1997, p. 469). The ability to affect a change in policy based on any discovered sectoral trends determines that a governmental entity is the appropriate perspective from which to approach this analysis.

The study by Lahr & Stevens (2002) deals with the historical rationale behind the use of aggregated data in regional analysis models. Data availability or the lack thereof, at the regional level, has traditionally been the reason for the misapplication of national level data in a regional analysis. The cost effectiveness of utilizing national Input-Output tables for use as regional weights is offset by the distortions introduced into the model. Lahr & Stevens (2002) caution that the use of the 86-sector Standard Industrial Classification 2-digit codes for national I-O weights will result in the failure to reflect technological differences between the national and regional level. Aggregation at this level also combines firms with various input distributions into one sector, with one weight. This aggregation bias is further compounded by the homogeneous technology assumption that further misrepresents the inter-industry associations by smoothing the variations. To avoid inducing bias and inflating the resulting multipliers, Lahr & Stevens (2002) propose that the Industrial Mix (IM) is regionalized before it is aggregated.

Utilizing the territorial approach determined above, the attributes of the given territory are examined to determine the role that they execute in the location decisions of firms. The neoclassical approach was incorporated by Arauzo-Carod (2008) to associate location decisions with profit-maximizing and cost-minimizing strategies, which require the consideration of agglomeration economies, land prices, wages, transportation costs and workers' skills. As the starting point, Arauzo-Carod (2008) uses the work of Marshall (1920) on agglomeration economies, the external economies which derive benefits from the spatial concentration of firms and employment through specific labor markets, technology spillovers, and supplier accessibility. Glaeser & Kerr (2008) use industry-employment share by city to estimate local industrial conditions; this accounting for city-industry employment shares captures the view stressed by Marshall, Arrow, and Romer that concentrated industrial centers achieve gains in increasing returns. This theory emphasizes the natural cost advantages located in specific regions for particular industries.

The fixed boundaries of the administrative territorial units are not considered to be closed economies. Fernandez (2005) noted that the classical shift-share approach focuses on the relation of the analyzed region to national changes but does not take into account the interactions between adjacent geographical units. These interactions would be more appropriately accounted for within a TTWA unit, but the noted instability compromises the reliability of that model. The differentiation between the resident statuses of those employed within the region is not accounted for, but there is significant variation in the results dependent upon this factor. Shuai (2010) describes the different effects to the county's economic structure dependent upon the status of the employment recipient: resident unemployed will cause an increase in consumption spending within the locality with no gains in the real estate tax base and no additional impact costs, immigrating workers will create an increase in both consumption spending and the real estate and property tax revenues with the additional impact costs, and commuting workers will generate little in sales revenue within the locality with no change to real estate tax revenues or imposed impact costs. Shuai (2010) states that individuals are more likely to consume in close proximity to their homes, and in following, population size is closely related to the revenues derived from sales and meals taxes.

The model developed by Fotopoulos (2005) specifies that employment growth is derived from output growth which increases the demand for further inputs. Fotopoulos (2005) concludes by cautioning that firm size heterogeneity should not be ignored, that local conditions matter more than regional economic structure and the latter are not symmetrical across sectors when it comes to the effects of business demographics on regional employment. The results from this business demographics shift-share analysis suggest that econometric models analyzing regional employment variations should control for differences in local conditions. To accomplish this, Kang (2007) employs a location quotient to calculate the presence of a cluster within the region. The concentration of an industry within the region is measured relative to the same industry's concentration at the national level.

The data selected for this model was evaluated for compatibility between the available reporting agencies. The data considerations are addressed by Fairman *et al.* (2008) in reporting on the key differences in the data provided by the Bureau of Labor Statistics and the Bureau of the Census. The results in her study indicated that within the same industrial sector, payroll values varied by as much as $95 billion between the two reports. Fairman *et al.* (2008) also identified the scope of the BLS Quarterly Census of Employment and Wages as encompassing more industries than the BOC. Furthermore, the BLS publishes its data quarterly, with a six-month lag and revises the data every quarter for the periods within the same calendar year. This constant revision improves the accuracy in the reported data whereas the BOC does not revise its County Business Patterns published data.

IV. Theoretical Model

Based on a Cobb-Douglas production function, this model assumes that *Labor*, *Capital*, and *Technology* are the inputs utilized in production. *Labor* is held to be the only input manipulated by any firm, and that *Capital* and *Technology* are fixed. To develop a model that relates changes in employment to changes in demand for the firm's goods, it is necessary to assume that *Labor* inputs are subject to diminishing marginal returns and that each firm demonstrates profit-maximizing behavior. A firm producing at the point where the marginal cost of production is equal to marginal revenue will adjust *Labor* to meet changes in demand. Holding *Capital* and *Technology* constant, as demand for the firm's goods increases, labor inputs will be increased to boost output. Utilizing this assumption, the observed changes in employment within an industry reflect changes in demand for the industry's goods. Adjusting the model's field of analysis to the industrial sector allows for an increase in sectoral employment to capture the resulting change from a stronger demand for that industry's output. Decreasing sectoral employment indicates weaker demand for the given industry's output. These variations in sectoral employment allow for inter-firm labor mobility and inter-industry labor transfers. There is no distinction in the quality of the labor inputs; all labor is viewed as homogenous within the given industry.

This study utilizes data provided by the Bureau of Labor Statistics to analyze changes in Campbell County's industrial composition. The BLS has developed a 5-digit industry classification system based on the primary production output of the firm, called the North American Industry Classification System (NAICS). This system was implemented in 1997 to replace the Standard Industry Classification (SIC). One of the limitations of using NAICS codes is the lack of compatibility with the SIC codes and any data utilizing that system. To remedy this compatibility problem and the imposed time constraints on any historical analysis, the BLS has transformed the archived SIC-based data to the NAICS format beginning with the year 1990. The NAICS industry codes, descriptions, and variable identifiers are listed in Table 1. Although 2-digit sector codes do not differentiate between the involved firms, the ability to examine industrial clustering is still present.

The Quarterly Census of Employment and Wages (QCEW) provides the number of establishments, number of employees, and average weekly wage, both as an aggregate for the study area and by industry classification for both Campbell County and the state of Virginia. The Labor Market Information division of the Virginia Employment Commission enables users to define queries for extracting area-specific data from the QCEW. The non-disclosed data utilized in this model was provided by LMI and is readily available to government agencies. The period covered by this analysis begins with the 1st quarter of 1990 and continues through the 1st quarter of 2010 (1990:1 - 2010:1).

The QCEW data provides the beginning variables:

$\boldsymbol{CCEMP_i} = Campbell\ County\ Employment\ in\ industry\ "i"$
$\boldsymbol{CCWKWAGE_i} = Campbell\ County\ Weekly\ Wage\ in\ industry\ "i"$ [in nominal terms]
$\boldsymbol{VAEMP_i} = Virginia\ Employment\ in\ industry\ "i"$
$\boldsymbol{VAWKWAGE_i} = Virginia\ Weekly\ Wage\ in\ industry\ "i"$ [in nominal terms]
$\boldsymbol{CCEMPTOT} = Campbell\ County\ Total\ Employment$
$\boldsymbol{VAEMP} = Virginia\ Total\ Employment$

The Virginia Department of Taxation provides taxable sales data on a quarterly basis for Campbell County, VA. This is utilized as the variable:

$\boldsymbol{CCTAX} = Campbell\ County\ Taxable\ Sales$ [in nominal terms]

The data provided is transformed into the necessary format through the following equations:

EQN 1.1 $\boldsymbol{CCQTRWAGE_i} = CCWKWAGE_i * 13$

The variable created in EQN 1.1 is Quarterly Wages in Campbell County for industry "*I*". This adjustment is necessary to develop a common time interval for all the QCEW data to maintain temporal consistency for all of the terms included in the model.

EQN 1.2 $\boldsymbol{CCPAYROLL_i} = CCQTRWAGE_i * CCEMP_i$

EQN 1.2 establishes an interaction term between wages and employment within the given industry. This term represents the total income accrued by the employees in industry "i," which can be considered as the funds available for expenditure from "i." This term also establishes a common unit, dollars, for the dependent variable (***CCTAX***) and the terms which represent each industry as independent variables. This common unit allows for the direct comparison of the dependent variable with all of the explanatory variables. The next series of equations are applied to generate an industry-specific location quotient:

EQN 2.1 $\boldsymbol{VALQ_i} = \frac{VAEMP_i}{VAEMP}$

The term generated by EQN 2.1 defines the employment share of industry "i" within Virginia's Total Employment. The concentration of the sectoral employment at the aggregate level is intended to represent the strength of the industry. Any fluctuations in this term are considered to be industry-wide fluctuations, or the performance metric for the sector. The same methodology utilized at the state level will be applied to the county level in EQN 2.2 to define the share of Campbell County's Total Employment that is located within Industry "i."

EQN 2.2 $\boldsymbol{CCLQ_i} = \frac{CCEMP_i}{CCEMPTOT}$

The concentration of sectoral employment at the regional level is a metric for local performance, influenced by local conditions. The terms from EQN 2.1 and EQN 2.2 are combined to quantify the employment density within each industry at the county level relative to the industry's concentration at the state level. This interaction produces the Employment-weighted Location Quotient for each NAICS classification.

EQN 2.3 $$\boldsymbol{LQ_i} = \frac{CCLQ_i}{VALQ_i} \quad \text{or} \quad \frac{CCEMP_i/CCEMPTOT}{VAEMP_i/VAEMP}$$

The term generated by EQN 2.3 is critical for the examination of the composition of the industrial base in a time-series model. This variable draws heavily from the two-period shift-share model where the change in industry "i" for a specific location is measured against change in the same industry for the entire region. If the $\boldsymbol{VALQ_i}$ is taken as a metric for sectoral performance and $\boldsymbol{CCLQ_i}$ is the regional performance metric, then $\boldsymbol{LQ_i}$ is the ratio of local industry strength relative to aggregate industry performance. The inclusion of $\boldsymbol{LQ_i}$ in the regression equation allows for the differentiation between aggregate industrial trends and local conditions (which are assumed to be the local advantages/disadvantages that will cause the local industrial performance to differ from aggregate performance). If the aggregate industry experiences a decline in employment while the local industry increases in employment, then it must be the local conditions that prevent the industry at the local level from following the trend of the aggregate industry. The dependent variable for the regression is ***CCTAX***, which is employed as a proxy for the economic activity within the region. To quantify the industrial contribution to the economic structure, the average quarterly wages for each industry are multiplied by the number of persons employed within that industry for the observed period. The location quotient is then applied as a weight to the industrial payroll to adjust for the suitability of the industry to the region.

The framework for each independent variable in the regression is specified by EQN 3.1.

EQN 3.1 $Z_i = CCPAYROLL_i * LQ_i$

The defined Z_i terms will represent the components of the industrial structure within the region. To this end, the sectoral relationships interacting in the model will describe the historical contributions of each NAICS 2-digit sector. The industrial agglomeration theory is extended beyond sectoral-clustering into inter-sectoral compatibility. For example in NAICS-11, *Firm A*'s primary product is an agricultural good, such as corn, soy, or tobacco. If this firm is successful, there is no indication that other NAICS-11 firms will migrate into the region as a result of *Firm A*'s success. This analysis of the industrial composition assumes that *Firm A* will need support services from other NAICS-designated sectors. Seed and fertilizer retailers are important supply firms for *Firm A,* as well as heavy equipment sales and service providers. Accessible bulk fuel distribution and transportation services are also required by *Firm A* to operate. This relationship is not unilateral; all of these firms cannot sustain themselves on the demand generated by *Firm A* alone. The presence of these firms within a region indicates that there is a market base strong enough to maintain their continued operation. This inter-industrial-dependence is expected to generate some level of multicollinearity between the independent variables representing each industrial sector. This multicollinearity will be more severe in interaction-intensive industries, such as transportation, since these industries have strong relationships with multiple sectors. Other industries are expected to fluctuate throughout the model in response to population changes, rather than the changes in a related sector. Retail Trade (44), Accommodation and Food Services (72), and Health Care and Social

Assistance (62) are all industries that are highly correlated with population and less so directly with other industrial sectors.

The development of the model and the specification of the variables are followed by the inspection of the input data. The initial examination of the dataset reveals some problems that will generate bias in the output estimation of the regression. The graphical representation of the data permits the clear recognition of outliers or problematic data that fail to represent the intended actual conditions. **Figure 1** shows that the weekly wages in NAICS industry 99 reflect an extreme outlier in comparison to the other industrial wages in the region, as well as the other observations within the same industry for the examined period. This observation, coupled with the negligible employment shares that NAICS 99 contributes to the region, necessitates that this variable be excluded from the model.

The employment trends during the observed period for the model follow an approximately linear path for each included industry. The problem presented in **Figure 2** is the observation for the 1st quarter of 2010 in NAICS 31. The industry experienced a 41 percent increase in employment, which is also captured in the Total County Employment term as a substantial gain. Further investigation into this sudden employment increase revealed that a firm previously claiming residency in an adjacent municipality had changed its residency status to Campbell County. This redesignation would be interpreted by the model as a substantial increase to regional employment without any of the structural changes that additional employment of this magnitude would affect. This gain to total employment will bias the $\boldsymbol{CCLQ_i}$ term by diminishing the regional concentration of all other industries for the observation period. Because there is no structural change in employment within the region, only a change in the reporting of data, the observation for the 1st quarter of 2010 is excluded from the model.

In Figure 3, the graph of weekly wages for NAICS industry 55 displays a level of variability that is not well suited to an OLS regression. Given that the observed functional form of this independent-variable-contributing term is non-linear, its exclusion from the model cannot be justified. NAICS 55 is a contributing sector to the industrial structure of the county; the elimination of this term will generate bias towards an inaccurate representation of the industrial composition of the county.

The exploration into industrial clustering begins with a dynamic shift-share analysis. The regional component of this analysis consists of Campbell County as a subset of the aggregate component, the state of Virginia. The relationship between the county and the state provides a higher level of commonality than the county has with any other territorial unit which possesses consistent and compatible data. This relationship serves to exclude some of the political differences that will mask the attributes of the area being examined.

The shift-share analysis is based on three components: the national share, the industry mix, and the regional shift. The first component, the national share (NS) represents the change in employment that would have occurred within industry *i* in Campbell County, if it had matched the rate of change in employment experienced at the state level.

(Equation 4.1) $$\boldsymbol{NS = CCEMP_{i_{[t-1]}} * \left[\frac{VAEMP_{[t]}}{VAEMP_{[t-1]}} - 1\right]}$$

(Where *i* refers to each industrial sector, *t* and *t-1* are the current and previous observation period respectively, and *CCEMP* and *VAEMP* are Campbell County and Virginia employment)

The second component of the shift-share analysis is the Industry Mix (IM) which determines the sectoral growth rate in employment at the state level relative to the aggregate growth rate in employment of the state. A positive IM indicates that industry growth at the state level exceeded net employment growth for the state aggregate. This term establishes the baseline for employment trends within each industrial sector. This employment trend is assumed to be the strength of the aggregate industry in excess of the net state employment fluctuations.

(Equation 4.2) $$IM = CCEMP_{i_{[t-1]}} * \left[\left(\frac{VAEMP_{i_{[t]}}}{VAEMP_{i_{[t-1]}}} - 1\right) - \left(\frac{VAEMP_{[t]}}{VAEMP_{[t-1]}} - 1\right)\right]$$

The final component in the shift share analysis is the Regional Shift (RS) which represents the local advantages that are attributed with generating the difference between the employment growth rates within the same industrial sector at the County and State levels. The RS term resulting in a positive value indicates an advantage within a given industrial sector at the regional level relative to the national level.

(Equation 4.3) $$RS = CCEMP_{i_{[t-1]}} * \left[\left(\frac{CCEMP_{i_{[t]}}}{CCEMP_{i_{[t-1]}}} - 1\right) - \left(\frac{VAEMP_{i[t]}}{VAEMP_{i_{[t-1]}}} - 1\right)\right]$$

The shift-share analysis is conducted on (18) NAICS industrial sectors. Several trends were revealed through examining the decomposition of employment changes. The National Share analysis reveals two noticeable trends when the growth rates are compared instead of changes in the levels of employment. Figure 4 depicts the pattern of seasonality observed in the calculated employment growth rates for the National Share.

Table 2.1 and Table 2.2 represent the highest and lowest growth rates respectively for the observed period. The first and third quarters are historically periods of decline, whereas the second and fourth quarters are periods of growth. This cyclical pattern is worth noting because of the effect that it will impose on the regression estimates, but no adjustment will be made to the data.

Figures 5 & 6 illustrate the seasonality present within the Regional Shift Component for NAICS 23 and NAICS 71. This seasonality affects the reliability of a two-period shift-share model.

The sectoral trends displayed in Table 3 are derived from a two-period shift share analysis. The Regional Shift (RS) is the component that this research will focus on. This component shows the change in employment that each sector experienced from the first quarter of 1990 to the fourth quarter of 2009. In Campbell County, industries **31**, **61**, **54** and **48** display the highest levels of decreased employment due to local conditions at (-6030), (-882), (-251), and (-250) respectively. Industries **23**, **62**, **51**, and **52** exhibit the highest levels of increased employment in the Regional Shift component at (1231), (430) (115) and (115) respectively. The two-period shift-share only utilizes the endpoints of the observed period and does not account for any of the changes that occurred in the interim. To ensure that all of fluctuations experienced by each sector are incorporated into the analysis, the average movement is calculated and displayed in Table 4.

The observed differences between the Regional Shift component and the Industry Mix component for each industry are examined to determine the sector's historical influence. Table 3 and Table 4 are compared to test the two-period based expectations against the dynamic expectations. The direction indicated by each table is consistent with the other, which justifies the usage of this analysis to determine the expected sign of the estimated beta coefficients in the regression model. Any industry that experienced a net gain in employment within the region while the established industry baseline experienced a loss (or a significantly smaller gain) is considered to have a regional advantage in that industry. Any industry that experienced a net loss in employment within the region while the established industry baseline experienced a gain is considered to have a regional disadvantage within that industry. The expected negative coefficients in the regression model are for industries: **11, 31, 48, 54, 55, 61, 72,** and **81.** The expected positive coefficients are in industries: **21, 22, 23, 44, 51, 52, 53, 56, 62, 71, and 92.**

<u>Equation 1</u>

$$CCTAX=\beta_0 + \beta_1 Z11+ \beta_2 Z21+ \beta_3 Z22+ \beta_4 Z23 + \beta_5 Z31+ \beta_6 Z44 + \beta_7 Z48+ \beta_8 Z51+\beta_9 Z52+ \beta_{10} Z53 + \beta_{11} Z54 + \beta_{12} Z55 + \beta_{13} Z56 + \beta_{14} Z61+ \beta_{15} Z62+ \beta_{16} Z71 + \beta_{17} Z72 + \beta_{18} Z81+ \beta_{19} Z92+\varepsilon$$

V. Presentation and Analysis

The variables developed in the previous section are uploaded into the EViews 4.1 statistical analysis program. The frequency of the data is set to quarterly for the range from 1990:1 to 2009:4. A Least Squares regression is utilized to produce the *Best Linear Unbiased Estimators* with statistical tests applied to the regression model to ensure that there are no violations of the classical assumptions which will invalidate the *BLUE* classification. The number of observations is large enough to allow (60) degrees of freedom with the necessary explanatory variables. The results of the regression are displayed in Table 5.

Table 5 – Equation 1: Regression Results

Dependent Variable: CCTAX
Method: Least Squares
Sample: 1990:1 2009:4
Included observations: 80

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-6715267.	11374557	-0.590376	0.5572
Z11	8.312375	5.965998	1.393292	0.1687
Z21	-15.33016	11.01716	-1.391480	0.1692
Z22	12.91508	4.260451	3.031387	0.0036
Z23	0.247836	0.271782	0.911893	0.3655
Z31	0.018511	0.020589	0.899096	0.3722
Z44	4.484771	2.103704	2.131845	0.0371
Z45	2.870665	1.029576	2.788202	0.0071
Z51	4.631227	12.85873	0.360162	0.7200
Z52	1.821724	1.438863	1.266086	0.2104
Z53	-161.1014	24.98192	-6.448720	0.0000
Z54	2.422124	4.583986	0.528388	0.5992
Z55	-9.110231	9.345845	-0.974789	0.3336
Z56	-0.186448	1.792981	-0.103988	0.9175
Z61	0.902145	0.966861	0.933066	0.3545
Z62	16.61666	5.687973	2.921368	0.0049
Z71	-17.74663	24.42618	-0.726541	0.4703
Z72	21.47762	8.450855	2.541473	0.0136
Z81	6.985707	6.205121	1.125797	0.2647
Z92	0.600109	6.721197	0.089286	0.9292

R-squared	0.935012	Mean dependent var	70221018
Adjusted R-squared	0.914432	S.D. dependent var	14063614
S.E. of regression	4113893.	Akaike info criterion	33.50996
Sum squared resid	1.02E+15	Schwarz criterion	34.10546
Log likelihood	-1320.398	F-statistic	45.43372
Durbin-Watson stat	2.100092	Prob(F-statistic)	0.000000

**The variables highlighted in green were expected negative signs.*
**The beta coefficients highlighted in red are the estimated negative terms.*
**The p-values highlighted in yellow are statistically significant at (α=.10) and the purple-accented p-values are the borderline cases.*

The relationship between the (19) explanatory variables and *CCTAX* displays (6) statistically significant variables with (2) borderline significant variables at the (α=.10) confidence interval for a two-tailed test. The estimated beta coefficients for the **Z21, Z53, Z55,** [redacted], and **Z71** variables indicate a negative influence exerted by these industries on the dependent variable. The expected signs derived from the shift-share analysis based on employment shares are consistent for **Z55.** The negative influence for **Z71** can be attributed to this industry consistently reporting the lowest wages in the county. The expected negative sign for **Z31** and **Z61** was not realized in the regression output, but the magnitude of these sectors' coefficients was substantially small. The **Z11, Z48, Z54, Z72,** and **Z81** terms were expected to induce a decrease in the dependent variable based on solely employment shares. This expectation was invalidated by adjusting employment shares to account for wage differences and incorporating the sectoral changes in a dynamic framework. The same effect on expectations for **Z21, Z53,** and **Z71** was achieved by wage inclusion and dynamic analysis. The wages for **Z53** are relatively low wages for the area and **Z71** offered the lowest in the county for the observed period. The wage adjustment to employment shares for these two industries accounts for the negative coefficient in the estimates.

The necessary tests have been conducted to insure that there have been no violations of the classical assumptions. These tests are displayed in **Appendix B** in <u>Table 6</u> through <u>Table 8</u>. With the exception of multicollinearity, which is expected in a model of this nature, these conditions have been satisfactorily met. The last adjustment to this model is made to adjust for the seasonality observed in the QCEW data and the location quotient for several of the industrial sectors. Three intercept dummy variables are introduced to correct for any bias that seasonality causes in <u>Equation 1</u>. The first quarter observation is the default and receives no dummy variable; a value of (1) is incorporated into the **X2**(2^{nd} qtr), **X3**(3^{rd}), or **X4**(4^{th} qtr) term if the observation period corresponds with that variable, a value of (0) is otherwise assigned.

<u>Equation 2 Regression</u>

$$CCTAX=\beta_0 + \beta_1 Z11+ \beta_2 Z21+ \beta_3 Z22+ \beta_4 Z23 - \beta_5 Z31+ \beta_6 Z44 + \beta_7 Z48+ \beta_8 Z51+\beta_9 Z52+ \beta_{10} Z53 + \beta_{11} Z54 + \beta_{12} Z55 + \beta_{13} Z56 + \beta_{14} Z61+ \beta_{15} Z62+ \beta_{16} Z71 + \beta_{17} Z72 + \beta_{18} Z81+ \beta_{19} Z92+ \beta_{20} X2 + \beta_{21} X3+ \beta_{22} X4+ \varepsilon$$

The justifications for different coefficient sign arrived at in the first regression equation are applied the <u>Equation 2</u>. Negative coefficients are expected for **Z31, Z53, Z55, Z61, & Z71.**

<u>Table 9 – Equation 2 Regression Results</u>

Dependent Variable: **CCTAX**
Method: Least Squares
Sample: 1990Q1 2009Q4
Included observations: 80

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-448532.4	11055502	-0.040571	0.9678
X2	5937890.	1561432.	3.802848	0.0004
X3	1215352.	2180326.	0.557418	0.5794
X4	4072930.	1937864.	2.101762	0.0400
Z11	10.62965	6.168094	1.723328	0.0903
Z21	-22.44533	10.06706	-2.229581	0.0297
Z22	14.44650	3.933319	3.672853	0.0005
Z23	0.075772	0.271652	0.278929	0.7813
Z31	-0.000322	0.021593	-0.014890	0.9882
Z44	3.698825	2.100362	1.761042	0.0836
Z48	2.264580	0.959536	2.360079	0.0217
Z51	14.85699	12.43997	1.194295	0.2373
Z52	1.785649	1.401268	1.274310	0.2077
Z53	-176.4570	23.32611	-7.564783	0.0000
Z54	5.594870	4.490938	1.245813	0.2179
Z55	-9.500436	8.560459	-1.109804	0.2717
Z56	-0.164653	1.660321	-0.099169	0.9214
Z61	2.782090	1.506427	1.846814	0.0700
Z62	14.12479	5.501127	2.567618	0.0129
Z71	-10.63125	22.53688	-0.471727	0.6389
Z72	17.94688	8.019399	2.237934	0.0291

Z81	0.603456	6.422820	0.093955	0.9255
Z92	-1.765762	6.278386	-0.281245	0.7795

R-squared	0.950076	Mean dependent var	70221018
Adjusted R-squared	0.930807	S.D. dependent var	14063614
S.E. of regression	3699383.	Akaike info criterion	33.32125
Sum squared resid	7.80E+14	Schwarz criterion	34.00609
Log likelihood	-1309.850	Hannan-Quinn criter.	33.59582
F-statistic	49.30580	Durbin-Watson stat	1.913034
Prob(F-statistic)	0.000000		

**The variables highlighted in green were expected negative signs.*
**The beta coefficients highlighted in red are the estimated negative terms.*
**The p-values highlighted in yellow are statistically significant at (α=.10)*

The addition of (3) dummy variables to adjust for seasonality in the observations improved the regression results. The addition of (3) independent variables enhanced the adjusted R-squared as well as a significant increase to the F-Statistic. The number of significant explanatory variables increased to (11) out of (22) at the (α=.10) level for a two-tailed test. A two-tailed test is still conducted because the expected sign of the coefficients is based on employment shares alone, and the expression of the industry terms is weighted in a manner that invalidates the use of a one-tailed test. The Akaike info Criterion (AIC) and the Schwarz Criterion (SC) were both lower in <u>Equation 2</u> relative to <u>Equation 1</u> which shows that the inclusion of the seasonal adjustment dummy variables improved the overall fit of the model, even with the higher assessed penalty in these calculations for additional explanatory variables. The joint hypothesis test for the overall significance of the model defined by <u>Equation 2</u> is conducted based on the generated F-Stat of (49.30580).

The Null Hypothesis:

$$H_0: \beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = \beta_6 = \beta_7 = \beta_8 = \beta_9 = \beta_{10} = \beta_{11} =$$

$$\beta_{12} = \beta_{13} = \beta_{14} = \beta_{15} = \beta_{16} = \beta_{17} = \beta_{18} = \beta_{19} = \beta_{20} = \beta_{21} = \beta_{22} = 0$$

The Alternative Hypothesis:

$H_A: H_0$ *is not true*

The interpolated critical F-statistic for (k = 23) and (d.f. = 57) is ($F_C \approx 1.62$) at the (α = .01) significance level. The magnitude of the calculated F-stat in excess of the critical-F leads to the rejection of the Null Hypothesis. Equation 2 is statistically significant at the 1% level.

The residual graph in Figure 8 illustrates the variance in the error term does not exhibit heteroskedasticity. The fluctuations from observation to observation in the residuals do not the follow the pattern of previous observations affecting the current value of the residuals, which would serve as evidence of positive serial correlation.

The Ramsey RESET test on Equation 2 (displayed in Table 10) indicates that there is no model misspecification by comparing the calculated F-Stat and the critical F-stat based on both the numerator and denominator's degrees of freedom. The null hypothesis in the RESET Test is that the beta coefficients of the fitted terms are all (0) and that no model misspecification exists. The critical-F is interpolated from at ($F_C \approx 2.78$) for (k=3) and (d.f. =57) at the 5% significance level. The calculated F-stat (0.792978) is below the critical-F (2.78) enough to justify not rejecting the Null Hypothesis. Equation 2 does not suffer from model misspecification or omitted variable bias.

The multicollinearity present in **Equation 1** is still evident in **Equation 2.** The VIFs were slightly increased for each variable, but there were no significant increases that could not be attributed to the addition of (3) explanatory variables. First-order serial correlation is tested for utilizing the Serial Correlation LM test. The ability of the lagged residual to influence the current observation's residual is evaluated by the Chi-square test. The p-value for the significance of the lagged residuals' influence on current residuals is highlighted in <u>Table 11</u>. The probability of (.683919) is not significant enough to justify the rejection of the null hypothesis of no first order serial correlation.

<u>Equation 2</u> produces a model that is statistically significant at the (α=.01) level. The adjusted R^2 of (0.930) indicates that 93% of the variation in taxable sales in Campbell County is explained by the industrial composition of the region. Of the (22) explanatory variables, (11) are significant at the (α=.10) for a two-tailed test. The industrial trend for each sector's variable indicates that **Z53, Z21, Z71,** and **Z55** have historically exerted a strong negative influence of the regional economy relative to the other industrial components. The sectoral trends estimated in this model also indicate that **Z72, Z51, Z22, Z62,** and **Z11** have exerted a strong positive influence on the regional economy relative to the sectors represented in the region.

I. Policy and limitations

This model is intended to assist local government in assessing trends in regional industrial composition and satisfies the stated requirements for ease of interpretation, low cost of generating the analysis, and the utilization of readily available, compatible data. This research has centered around identifying the industries that have exerted a negative influence on the region. Rather than developing a model that dictates which industries to target for marketing, this model identifies those industries that are not well suited to the region. This identification preserves the discretion that economic development professionals must exercise in the performance of their duties. Focusing on the best-suited or best-performing industries limits the scope of recruiting opportunities. A well-suited industry should not be ignored because it has not historically performed as the best-suited industry or high as another top industry. The identification of those industries that have negatively influenced the county's economy is extremely valuable.

The regression model results display industrial trends with beta coefficients that are directly comparable to one another. A rank order can be established from the best-suited to the worst-suited industries for the region. Industries 72, 22, and 62 are identified by the model as the sectors which have historically exerted the strongest positive influence on the regional economy. Industries 53, 21, and 71 are the strongest negative sectoral influences. The relationship between industry trends and taxable sales becomes difficult to explain in terms of increased employment within an industry causing a decrease in economic activity. The use of taxable sales as the dependent variable makes the regression-estimated industrial trends difficult to justify. The most helpful information for policy considerations comes from the shift-share analysis.

The shift-share analysis results for Campbell County employment changes offer some useful insights. A negative Industry Mix paired with a negative Regional Shift for a given industry indicates a declining industry with poor regional performance. This trend has been observed in 31, 44, and 48; all of which have experienced a regional decline in excess of the aggregate industrial decline. Industries 21, 22, 23, 51, 52, 53 and 92 all experienced a positive Regional Shift and a negative Industry Mix. This relationship can indicate that Campbell County has locational advantages that allow these industries to perform better in the region, or that the decline experienced by the aggregate industry has not yet affected the regional industries. In industries 71 and 56, the IM and RS are both positive, but the change in IM is twice the growth observed in the RS component. The positive growth in the RS in excess of the positive growth in the IM only occurs in industry 62. This indicates a growing industry that is well-suited to the region.

The expression of each industry's contribution in this model does not distinguish between changes in employment and changes in wages. There is a fundamental limitation that this expression imposes on the model. It is evident that a given region displays wage contours, which would be the wage differentials between industries or the regional average. Any change in this system of contours would cause a ripple effect throughout all of the wage levels. A firm entering the region that offered a substantially higher wage for employees than was present in the system would bid up the purchasing price of labor inputs. All firms with comparable quality labor would have to increase wages to retain their employees or lose those inputs to higher wage opportunities. The model developed in this paper is not capable of capturing this interaction between firms competing for labor or the differences in wages between firms. The location quotient employed in this model is utilized to capture the changes in relative sectoral concentration but is not sufficient for incorporating the wage structure- induced changes to industrial composition.

There are some precision limitations to take into consideration before any policy can be formulated based on these results. Spatial aggregation distorts the local conditions. The population, employment opportunities, and retail opportunities are assumed by the model to be uniformly distributed throughout the geographic region. In reality, the varying density of these regional attributes is clustered in pockets throughout the area. Site selection within the region is important to connect sectoral advantages with the appropriate firm.

The metric used to relate employment to taxable sales was wages in this model. Relying on wages, instead of total compensation, distorts the relationship that various firms have with the region. Total compensation, to include all employer-provided benefits, would greatly enhance the precision of the explained relationships in the model. The consumption patterns of the population are also not accurately represented in this model. There is no differentiation between the spending patterns of the various income ranges, and only income from employment is utilized. The available funds in the region that were not generated by employment are completely discounted in this model.

The development of this model implies that income generated in a given quarter is also spent during the same period. A better formulation would include some level of taxable sales from the previous period as an explanatory variable. As an alternative to a simple lag or a moving average, the exponential smoothing of the taxable sales should be examined. This process would be preferable because the recent observations are assigned a higher weight that decreases as the observations extend farther back in time. This additional explanatory variable would certainly introduce serial correlation into the model, but future research should focus on incorporating some form of a regressive lagged dependent variable as an independent variable that does not compromise the statistical integrity of the model.

VI. References

Arauzo-Carod, J.-M. (2008). Industrial Location at a Local Level: Comments on the Territorial Level of the Analysis. *Tijdschrift voor Economische en Sociale Geografie , Vol. 99* (No. 2), 193-208.

Bureau of Labor Statistics. (2011, Feb 17). *Consumer Price Index, All Urban Consumers.* Retrieved March 6, 2011, from United States Department of Labor - Bureau of Labor Statistics: ftp://ftp.bls.gov/pub/special.requests/cpi/cpiai.txt

D'Elia, J. L. (2005). *Labour Productivity, Gross Value Added and Employment by Industry in Northern Ireland. a Structural and Shift-Share Analysis.* Belfast: Economic Research Institute of Northern Ireland.

Dinc, M., & Haynes, K. E. (1997). Sources of Regional Inefficiency: An Integrated shift-share, data envelopment analysis and input-output approach. *The Annals of Regional Science , VOL. 33*, 469-489.

Fairman, K., Foster, L., Krizan, C., & Rucker, I. (2008). *An Analysis of the Key Differences in Micro Data: Results from the Business List Comparison Project.* Washington, DC: Center for Economic Studies, Bureau of the Census.

Fernandez, M. M. (2005). Spatial shift-share analysis: new developments and some findings for the Spanish case. *45th Congress of the European Regional Science Association* . University of Oviedo, Department of Applied Economics.

Fotopoulos, G. (2005). A Business-Demographics Adjusted Shift-Share Analysis: The effects of business demography on regional employment and output growth. *Papers on Entrepreneurship, Growth and Public Policy* . Max Planck Institute of Economics, Entrepreneurship, Growth and Public Policy Group.

Fotopoulos, G. (2007). Integrating Firm Dynamics into the Shift-Share Framework. *VOL. 38(1)*, 140-152.

Gabe, T. M. (2006). Growth of Creative Occupations in U.S. Metropolitan Areas: A Shift-Share Analysis. *Growth and Change , Vol. 37* (3), 396-415.

Glaeser, E. L., & Kerr, W. R. (2008). *Local Industrial Conditions and Entrepreneurship: How much of the Spatial Distribution Can We Explain?* Washington, DC: Center for Economic Studies, Bureau of the Census.

Guimaraes, P., Figueiredo, O., & Woodward, D. (2004). *Location Modeling and the Localization of Portuguese Manufacturing Industries.* European Regional Science Association.

Gujarati, D. N. (2006). *Essentials od econometrics* (3rd ed.). New York: McGraw-Hill Irwin.

Kang, Y. (2007). Made in China: Coastal Industrial Clusters and Regional Growth. *Issues in Political Economy* .

Labor Market Statistics, Covered Employment and Wages Program. (2010, Sept 14). *Virginia Workforce Connection - Quarterly Census of Employment and Wages - Multiple Quarterly time periods for Quarterly Census of Employment and Wages, Sector (2 digit) data for Campbell County, VA, Aggregate of all types.* Retrieved Sept 14, 2010, from Virginia Employment commission: http://www.vawc.virginia.gov/analyzer/qsInd202.asp?quicksearch=True&setvar=True&cat=HST_EMP_WAGE_IND&session=IND202&subsession=99&areaname=

Labor Market Statistics, Covered Employment and Wages Program. (2010, Sept 14). *Virginia Workforce Connection - Quarterly Census of Employment and Wages - Multiple Quarterly time periods for Quarterly Census of Employment and Wages, Sector (2 digit) data for Virginia Statewide, Aggregate of all types.* Retrieved Sept 14, 2010, from Virginia Employment Commission: http://www.vawc.virginia.gov/analyzer/qsInd202.asp?quicksearch=True&setvar=True&cat=HST_EMP_WAGE_IND&session=IND202&subsession=99&areaname=

Lahr, M. L., & Stevens, B. H. (2002). A Study of the Role of Regionalization in the Generation of Aggregation Error in Regional Input-Output Models. *Journal of Regional Science , VOL. 42*, pp. 477-507.

Martins, E. (1993). *Construction of Regional Input-Output Tables from Establishment-Level Microdata: Illinois, 1982.* Washington, DC: Center for Economic Studies, Bureau of the Census.

Shuai, X. (2010, January). Who Benefits from Job Creation at County Level? An Analysis of Leakage and Spillover of New Employment Opportunities in Virginia. *Business Economics , VOL. 45(1)*, pp. 38-48.

Studendmund, A. (2006). *Using Econometrics: A Practical Guide* (5th ed.). Boston: Pearson Addison-Wesley.

Virginia Department of Taxation. (2010, Sept 14). *Taxable Sales Reports - Taxable Sales by quarter, localities, 1984-2010 (1st Quarter).* Retrieved Sept 2010, 2010, from Weldon CooperCenter for Public Service: http://www.coopercenter.org/econ/taxablesales

Appendix

Table 1 NAICS 2-Digit Codes

Variable	NAICS	Industry Title
11	11	Agriculture, Forestry, Fishing and Hunting
21	21	Mining
22	22	Utilities
23	23	Construction
31	31-33	Manufacturing
42	42	Wholesale Trade
44	44-45	Retail Trade
48	48-49	Transportation and Warehousing
51	51	Information
52	52	Finance and Insurance
53	53	Real Estate and Rental and Leasing
54	54	Professional, Scientific, and Technical Services
55	55	Management of Companies and Enterprises
56	56	Administrative and Support and Waste Management and Remediation Services
61	61	Education Services
62	62	Health Care and Social Assistance
71	71	Arts, Entertainment, and Recreation
72	72	Accommodation and Food Services
81	81	Other Services (except Public Administration)
92	92	Public Administration
99	99	Unclassified

Figure 1 – Campbell County Weekly Wages by NAICS

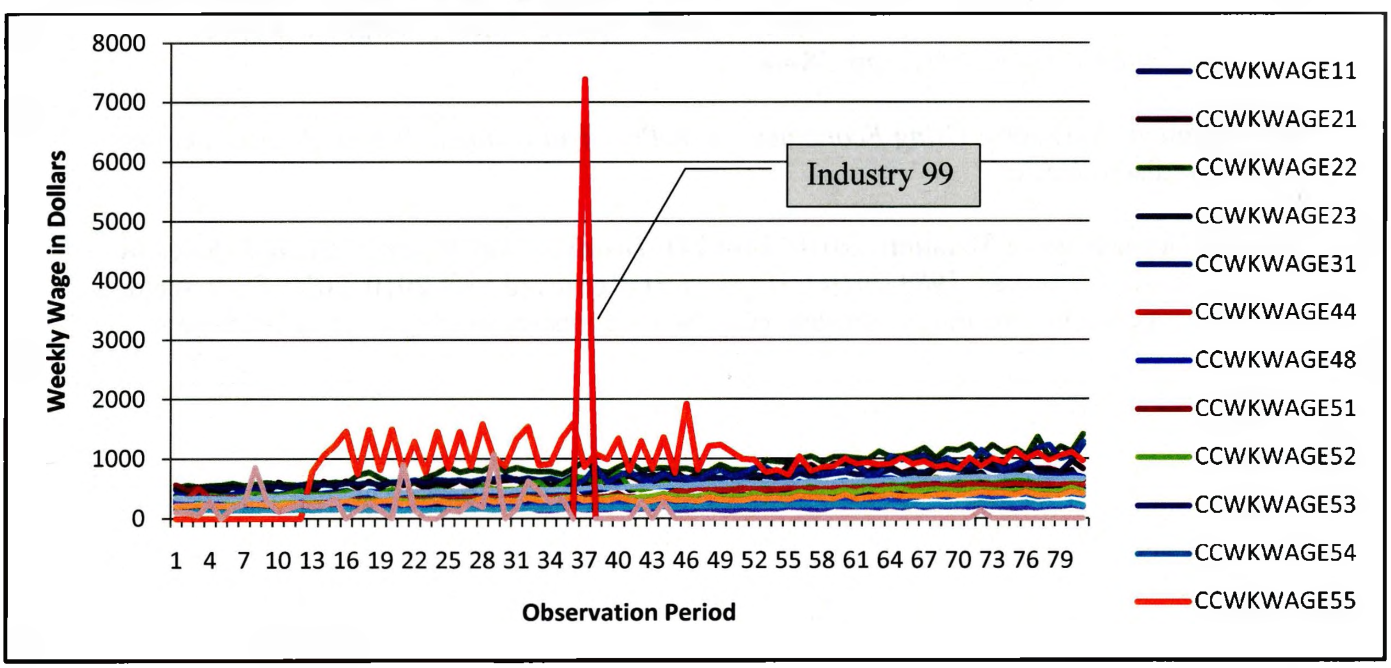


Figure 2 – Campbell County Employment by NAICS

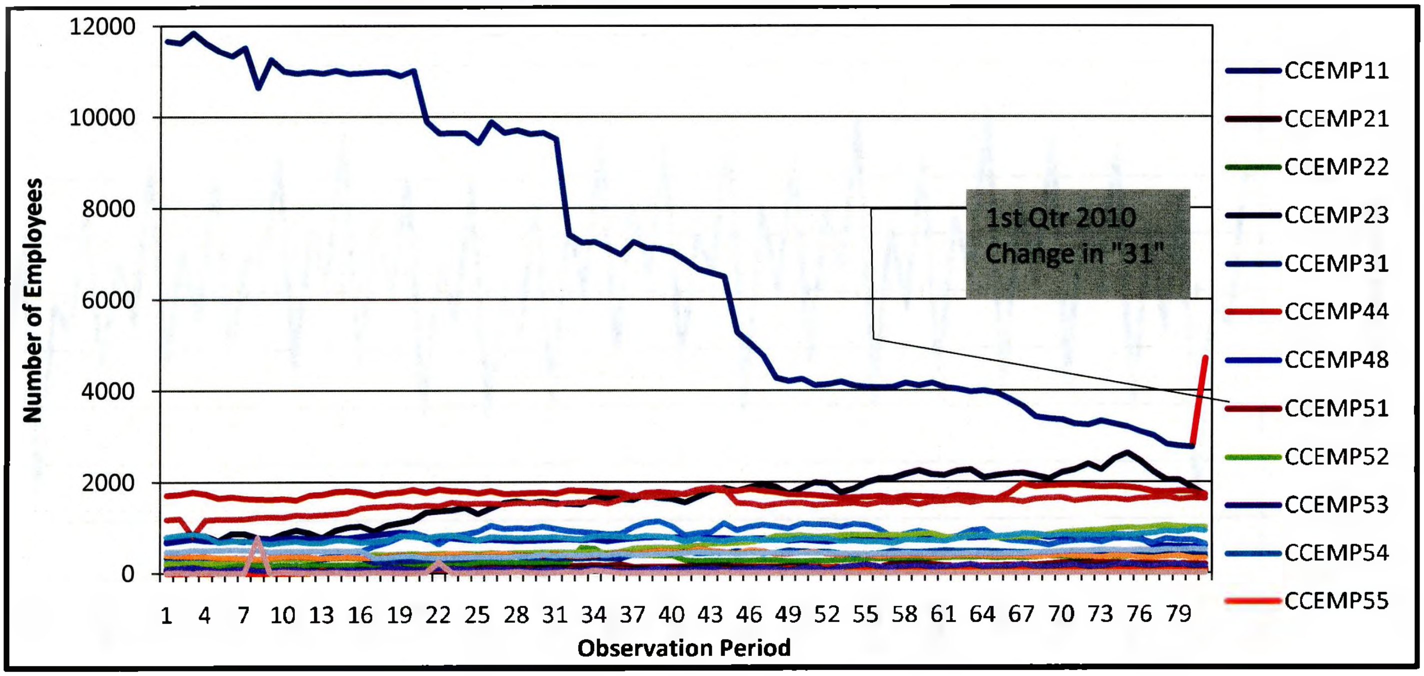


Figure 3 – Campbell County Weekly Wages for NAICS industry 55

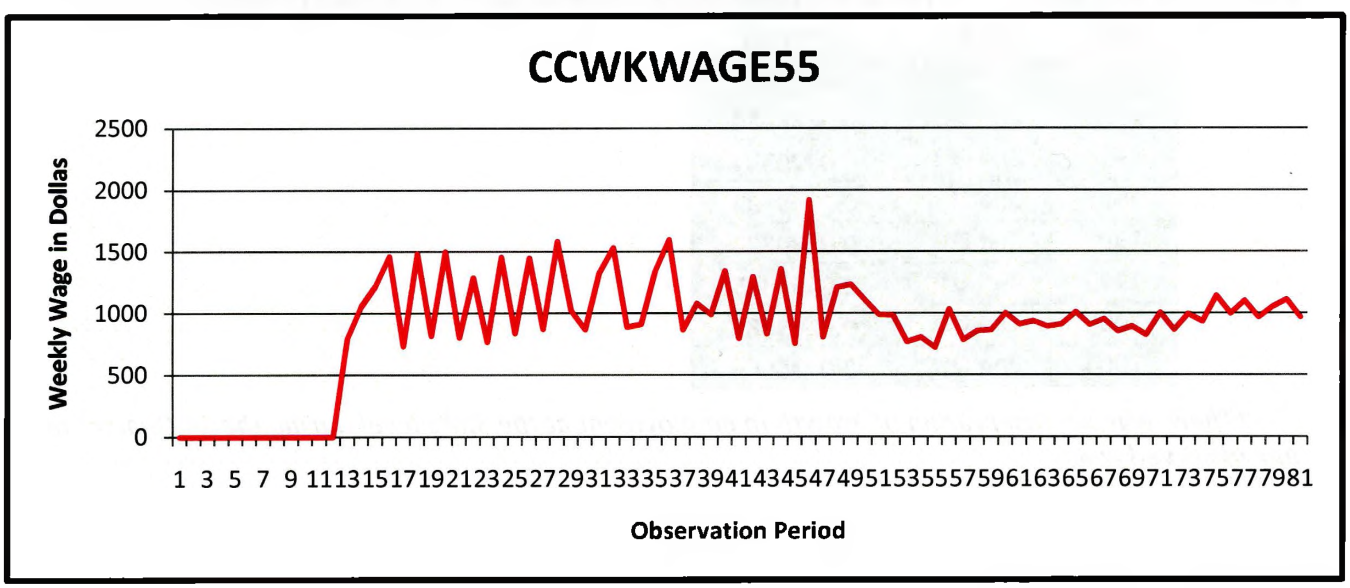


Figure 4 – Quarterly Growth Rate of National Share (NS) in Employment

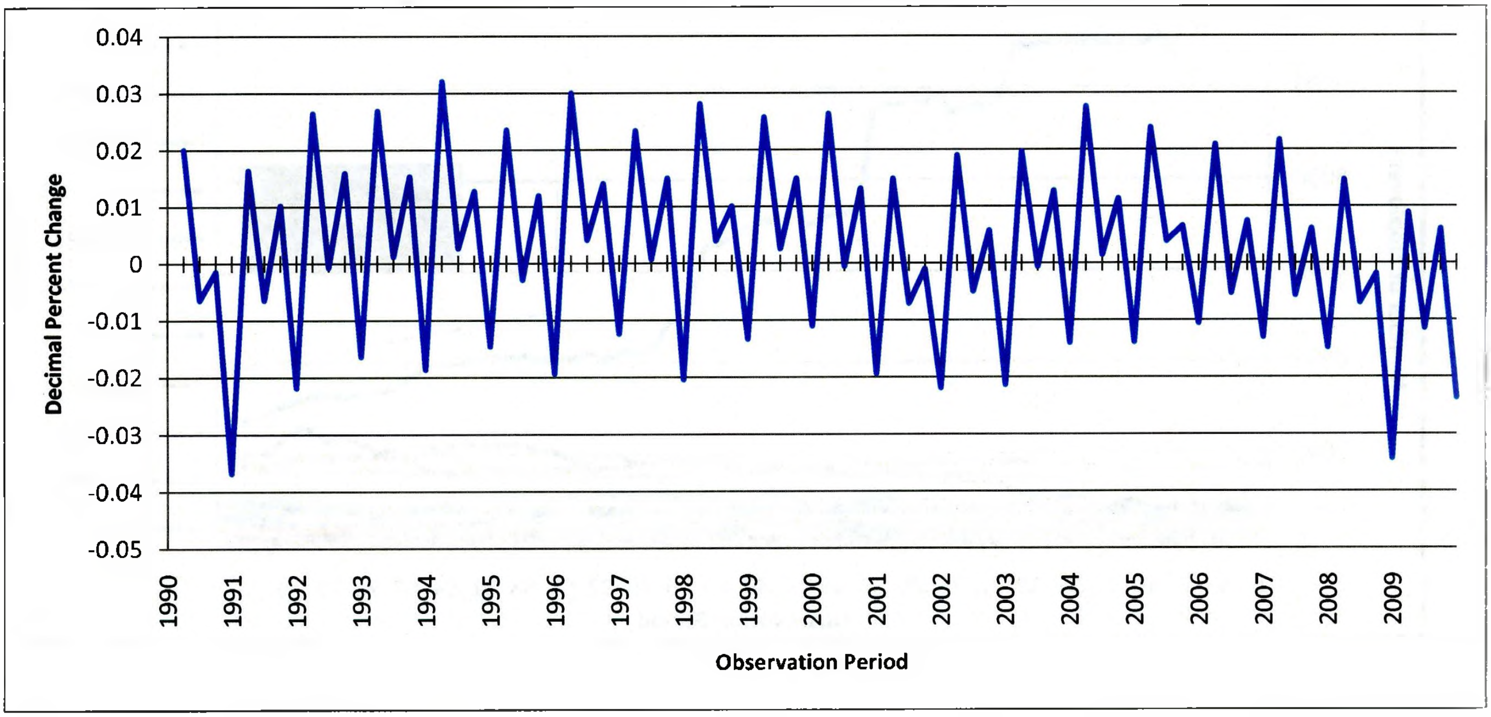


Table 2.1

Shift-Share Analysis: National Share Highest Growth Periods in Virginia Employment

Year	Period	NS%CHG
1992	2nd Qtr.	0.0264684
1993	2nd Qtr.	0.0268818
1994	2nd Qtr.	0.0320304
1996	2nd Qtr.	0.0299992
1998	2nd Qtr.	0.0280723
1999	2nd Qtr.	0.0256885
2000	2nd Qtr.	0.0263381
2004	2nd Qtr.	0.0275566

**There was no observation of growth in employment at the State level during the 1st Quarter of any observed year*

Table 2.2
Shift-Share Analysis: National Share Lowest Growth Periods in Virginia Employment

Year	Period	NS%CHG
1991	1st Qtr.	-0.0367668
1992	1st Qtr.	-0.0218238
1998	1st Qtr.	-0.0204235
2001	1st Qtr.	-0.0194479
2002	1st Qtr.	-0.0219511
2003	1st Qtr.	-0.0213928
2009	1st Qtr.	-0.0343186
2010	1st Qtr.	-0.0237018

**There was no observation of reduction in employment at the state level during the 2nd Quarter of any observed year*

Figure 5 –Quarterly Growth Rate in Employment for Regional Shift in 23

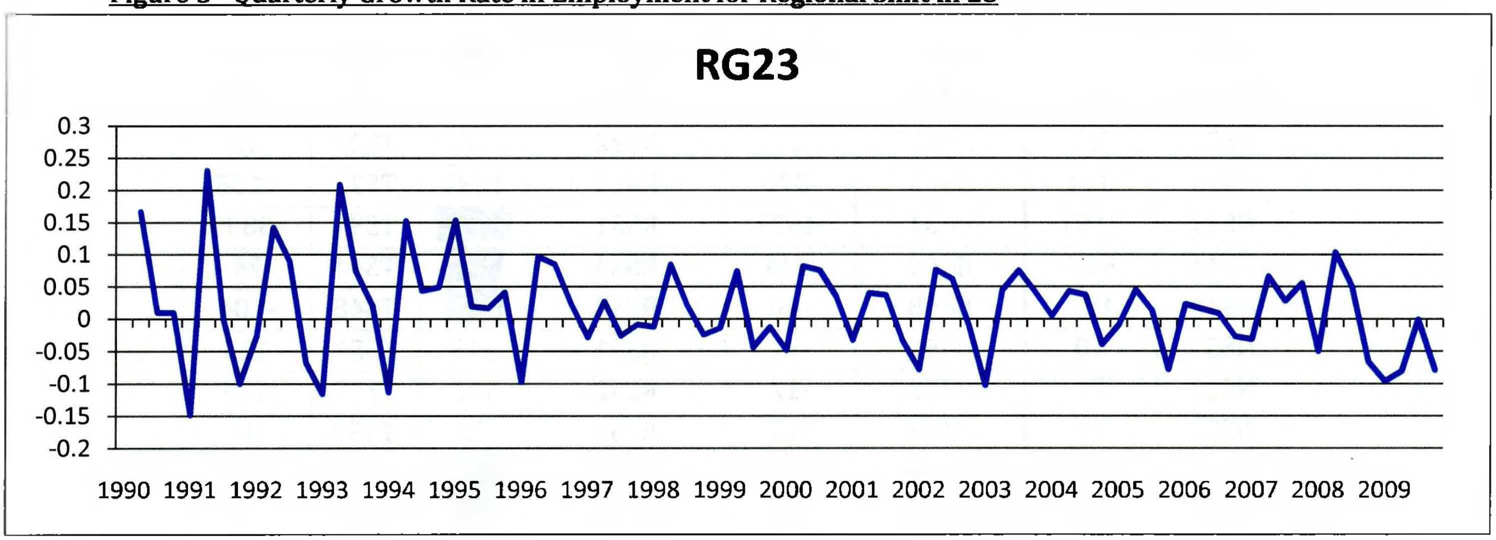


Figure 6 –Quarterly Growth Rate in Employment for Regional Shift in 71

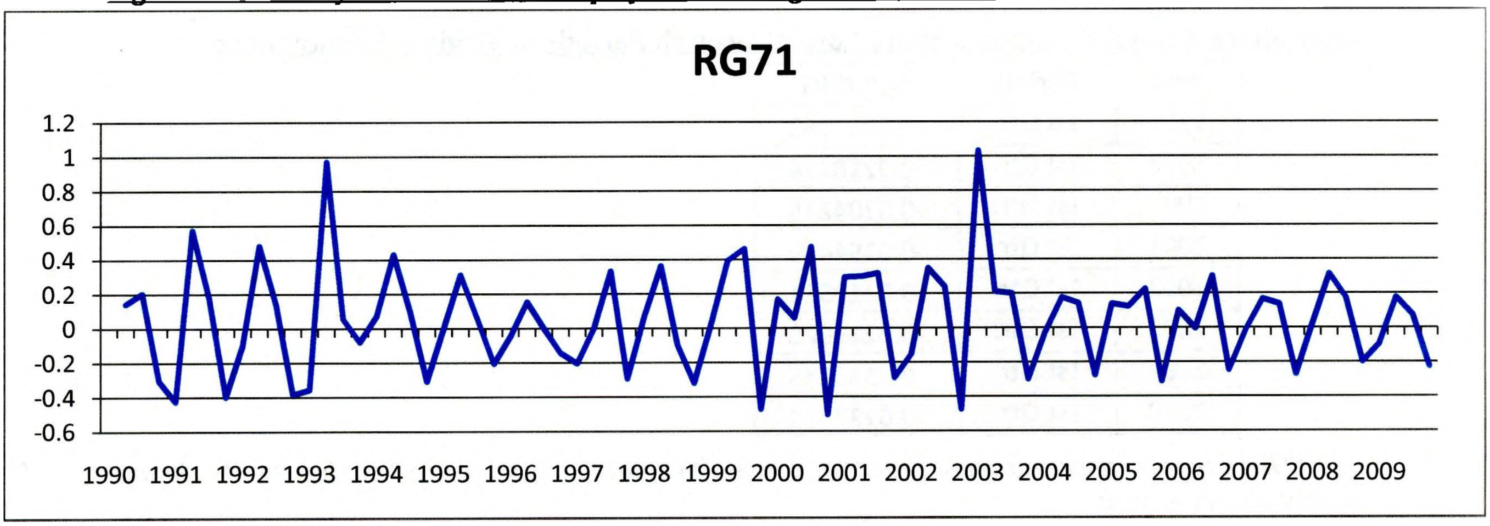


Table 3 - Shift Share Analysis Summed Totals By Component

1990:1-2009:4							
National Share Net Employment Change		Industry Mix Net Employment Change		Regional Shift Net Employment Difference		Net Total Shift	
NS11	21	IM11	77	RS11	[redacted]	TS11	23
NS21	13	IM21	-41	RS21	19	TS21	-10
NS22	23	IM22	-27	RS22	38	TS22	35
NS23	324	IM23	-370	RS23	1231	TS23	1185
NS31	2107	IM31	-4977	RS31	[redacted]	TS31	-8899
NS44	442	IM44	-144	RS44	[redacted]	TS44	74
NS48	187	IM48	-40	RS48	[redacted]	TS48	-104
NS51	40	IM51	-31	RS51	115	TS51	123
NS52	71	IM52	-12	RS52	115	TS52	175
NS53	20	IM53	-12	RS53	77	TS53	86
NS54	94	IM54	209	RS54	[redacted]	TS54	51
NS55	20	IM55	23	RS55	[redacted]	TS55	-5
NS56	187	IM56	204	RS56	95	TS56	485
NS61	374	IM61	976	RS61	[redacted]	TS61	468
NS62	120	IM62	236	RS62	430	TS62	786
NS71	24	IM71	18	RS71	9	TS71	51
NS72	172	IM72	149	RS72	[redacted]	TS72	156
NS81	96	IM81	11	RS81	[redacted]	TS81	-65
NS92	92	IM92	-88	RS92	56	TS92	60

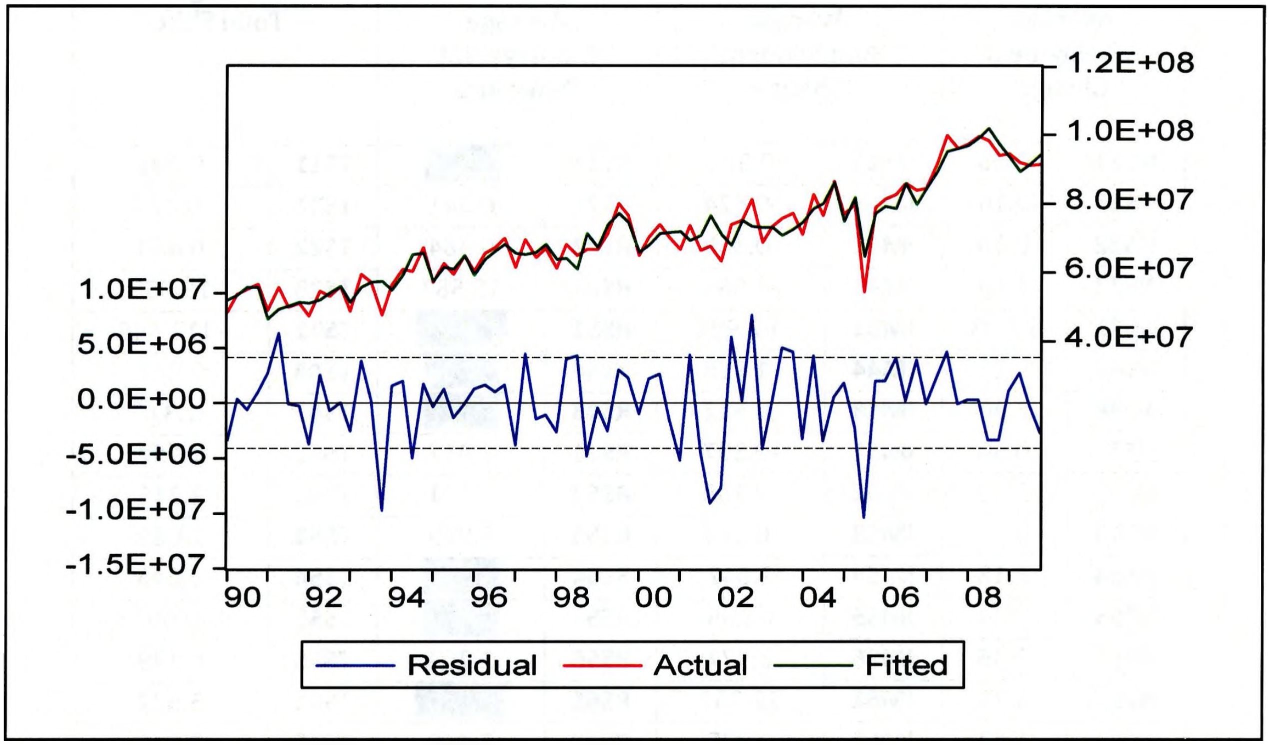
Table 4- Shift Share Analysis Average Change By Component

1990:1-2009:4							
National Share Average Employment Change		Industry Mix Average Employment Change		Regional Shift Average Employment Difference		Average Total Shift	
NS11	0.26	**IM11**	0.978	**RS11**	-0.946	**TS11**	0.291
NS21	0.16	**IM21**	-0.524	**RS21**	0.241	**TS21**	-0.127
NS22	0.30	**IM22**	-0.339	**RS22**	0.484	**TS22**	0.443
NS23	4.10	**IM23**	-4.689	**RS23**	15.587	**TS23**	15.000
NS31	26.68	**IM31**	-62.998	**RS31**	-76.328	**TS31**	-112.646
NS44	5.21	**IM44**	-1.818	**RS44**	-2.457	**TS44**	0.937
NS48	2.36	**IM48**	-0.512	**RS48**	-3.167	**TS48**	-1.316
NS51	0.50	**IM51**	-0.397	**RS51**	1.457	**TS51**	1.557
NS52	0.90	**IM52**	-0.148	**RS52**	1.461	**TS52**	2.215
NS53	0.26	**IM53**	-0.148	**RS53**	0.979	**TS53**	1.089
NS54	1.18	**IM54**	2.642	**RS54**	-3.174	**TS54**	0.646
NS55	0.24	**IM55**	0.295	**RS55**	-0.601	**TS55**	-0.063
NS56	2.36	**IM56**	2.577	**RS56**	1.203	**TS56**	6.139
NS61	4.73	**IM61**	12.351	**RS61**	-11.161	**TS61**	5.924
NS62	1.52	**IM62**	2.985	**RS62**	5.449	**TS62**	9.949
NS71	0.30	**IM71**	0.226	**RS71**	0.118	**TS71**	0.646
NS82	2.17	**IM72**	1.883	**RS72**	-2.083	**TS72**	1.975
NS81	1.22	**IM81**	0.134	**RS81**	-2.175	**TS81**	-0.823
NS92	1.16	**IM92**	-1.108	**RS92**	0.708	**TS92**	0.759

Appendix B

To verify the reliability of this model, the following tests are conducted to uncover any violation of the classical assumptions. In Figure 7, the residual graph shows no evidence of increased variation in the regression residuals across the observed period, which is indicative of homoskedasticity. The graph also shows no first-order perfect serial correlation in the error term in that the current observation of the error term is not a function of the previous observation of the residual.

Figure 7 – Residual graph



To test for impure serial correlation, the Durbin-Watson *d* statistic generated by the regression is evaluated against the reference value of (2). The close proximity of the calculated value to the reference value does not preclude the possibility of positive serial correlation for the given sample size and number of independent variables. The appropriate hypothesis test for positive serial correlation at the (α=.025) level states that:

$$H_0: p \leq 0, \text{ then no positive serial correlation}$$
$$H_A: p > 0, \text{ then positive serial correlation is present}$$

The calculated $\boldsymbol{d}$(2.100092) is between the $\boldsymbol{d_L}$(1.076) and $\boldsymbol{d_U}$(2.275) for $\boldsymbol{N} = 80, \boldsymbol{k} = 20$. Within this range, the test is inconclusive for positive serial correlation; the null hypothesis can neither be accepted nor rejected. The inability of the Durbin-Watson test to reject or not reject the hypothesis of no positive first order serial correlation leads to conducting the Ramsey RESET to verify the model is correctly specified.

Table 6 – Equation 1: Ramsey RESET Results

Ramsey RESET Test
Equation: UNTITLED
Specification: CCTAX C Z11 Z21 Z22 Z23 Z31 Z44 Z48 Z51 Z52 Z53 Z54 Z55 Z56 Z61 Z62 Z71 Z72 Z81 Z92
Omitted Variables: Powers of fitted values from 2 to 4

	Value	df	Probability
F-statistic	0.951005	(3, 57)	0.4222
Likelihood ratio	3.907244	3	0.2717

F-test summary:

	Sum of Sq.	df	Mean Squares
Test SSR	4.84E+13	3	1.61E+13
Restricted SSR	1.02E+15	60	1.69E+13
Unrestricted SSR	9.67E+14	57	1.70E+13
Unrestricted SSR	9.67E+14	57	1.70E+13

LR test summary:

	Value	df
Restricted LogL	-1320.398	60
Unrestricted LogL	-1318.445	57

Unrestricted Test Equation:
Dependent Variable: CCTAX
Method: Least Squares
Sample: 1990Q1 2009Q4
Included observations: 80

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-47063048	5.93E+08	-0.079424	0.9370
Z11	37.48340	206.1368	0.181838	0.8564
Z21	-64.42407	380.7827	-0.169189	0.8662
Z22	53.75386	320.2949	0.167826	0.8673
Z23	1.170625	6.222675	0.188123	0.8514
Z31	0.075787	0.458206	0.165398	0.8692
Z44	18.17544	110.8907	0.163904	0.8704
Z48	12.49208	71.33174	0.175127	0.8616
Z51	18.30509	114.1560	0.160352	0.8732
Z52	7.578301	45.22034	0.167586	0.8675
Z53	-685.2530	3997.322	-0.171428	0.8645
Z54	8.747657	61.08078	0.143215	0.8866
Z55	-34.83323	224.9982	-0.154816	0.8775
Z56	-0.079068	4.595299	-0.017206	0.9863
Z61	4.042786	22.42416	0.180287	0.8576
Z62	68.48478	411.1717	0.166560	0.8683
Z71	-71.62234	439.4385	-0.162986	0.8711
Z72	90.81438	531.2252	0.170953	0.8649
Z81	29.62826	173.7594	0.170513	0.8652
Z92	4.266220	17.63001	0.241986	0.8097
FITTED^2	-1.15E-07	5.30E-07	-0.216224	0.8296
FITTED^3	1.47E-15	4.92E-15	0.298618	0.7663
FITTED^4	-6.33E-24	1.68E-23	-0.376861	0.7077

R-squared	0.938109	Mean dependent var	70221018
Adjusted R-squared	0.914222	S.D. dependent var	14063614
S.E. of regression	4118942.	Akaike info criterion	33.53612
Sum squared resid	9.67E+14	Schwarz criterion	34.22095
Log likelihood	-1318.445	Hannan-Quinn criter.	33.81068
F-statistic	39.27177	Durbin-Watson stat	2.201752
Prob(F-statistic)	0.000000		

The (3) polynomial terms are included in the equation to act as proxies for any potentially omitted variables or misspecified functional form. The F-Statistic (0.951005) in the RESET test has a p-value far enough outside of the required range to not reject the null hypothesis of no model misspecification. With the possibility of an omitted variable eliminated, the likelihood of first order positive serial correlation resulting from omitted variable bias is also discounted.

The next violation examined is the presence of multicollinearity between the included independent variables. The first test is the simple correlation matrix with the stated correlation value of ($\pm$ 0.7) extracted. Table 7 indicates that several of the explanatory variables are highly correlated with one another.

Table 7 – Simple Correlation Matrix

Z11		**Z21**		**Z22**		**Z23**		**Z31**		**Z44**		**Z51**	
Z51	0.78133	**Z22**	0.805664	**Z23**	0.774034	**Z31**	-0.87568	**Z44**	-0.79584	**Z51**	0.739133	**Z53**	0.729842
		Z31	-0.74243	**Z31**	-0.76782	**Z44**	0.915858	**Z53**	-0.85061	**Z53**	0.918732	**Z62**	0.707042
		Z54	0.732183	**Z44**	0.716854	**Z51**	0.732016	**Z54**	-0.82946	**Z61**	0.822751	**Z92**	0.731485
		Z61	0.722195	**Z53**	0.754727	**Z61**	0.843245	**Z61**	-0.82607	**Z62**	0.868281		
				Z56	0.705027	**Z62**	0.947822	**Z62**	-0.91551	**Z72**	0.753478		
				Z61	0.81923	**Z72**	0.731809	**Z92**	-0.85594	**Z92**	0.851797		
				Z62	0.728144	**Z92**	0.895163						
Z53		**Z54**		**Z56**		**Z61**		**Z62**		**Z72**			
Z61	0.81689	**Z62**	0.731962	**Z81**	0.711558	**Z62**	0.804293	**Z72**	0.700823	**Z92**	0.748671		
Z62	0.92586					**Z92**	0.742355	**Z92**	0.913259				
Z72	0.7697												
Z92	0.8841												

The strong simple correlations present in the explanatory variables of the model justify the calculation of the Variance Inflation Factors for each independent variable. This test is calculated using EViews 7. The reference value for this test is (5) with any VIF in excess of this level indicating the presence of severe multicollinearity.

<u>Table 8 – Equation 1: Variance Inflation Factors</u>

Variance Inflation Factors
Sample: 1990Q1 2009Q4
Included observations: 80

Variable	Coefficient Variance	Uncentered VIF	Centered VIF
C	1.29E+14	611.5795	NA
Z11	35.59314	33.46474	6.606628
Z21	121.3778	56.62136	6.240167
Z22	18.15144	55.62543	8.803769
[redacted]	0.073865	91.26852	25.259[redacted]
[redacted]	0.000424	87.01187	24.094[redacted]
Z44	4.425572	487.8183	15.67412
Z48	1.060026	110.4136	3.946545
Z51	165.3471	55.03484	18.03123
Z52	2.070326	8.585620	2.891797
Z53	624.0965	54.96566	17.07430
Z54	21.01293	53.74598	10.45289
Z55	87.34482	14.09624	3.602793
Z56	3.214779	47.64728	9.702948
Z61	0.934820	196.3140	9.009550
[redacted]	32.35304	204.3660	71.22[redacted]
Z71	596.6381	25.97492	7.358169
Z72	71.41696	236.9428	9.838174
Z81	38.50353	111.4321	5.478347
Z92	45.17448	175.6122	16.44477

The VIFs indicate that only (3) of the explanatory variables do not exhibit severe multicollinearity, but the critical value of (5) may not be appropriate for a model with (19) explanatory variables. What is clear from this test is that the VIFs for **Z23, Z31,** and **Z62** exhibit extreme multicollinearity. Severe multicollinearity was expected for **Z62** because of the nature of that industry's relationship with population. The value for **Z31** can be intuitively understood by examining the drastic decline in that sector over the observed period. The change in employment for this industry is expected to be absorbed by the other sectors in the region. The negative correlation that **Z31** exhibited with the other sectors in <u>Table 7</u> would indicate that these sectors absorbed the employment shed by

Manufacturing. The high VIF for **Z23** and the negative simple correlation with **Z31**is a strong indication of the compatibility for the labor inputs between these two industries. No action is taken to correct for this violation of multicollinearity between the explanatory variables.

Figure 8 – Equation 2 Regression Residual Graph

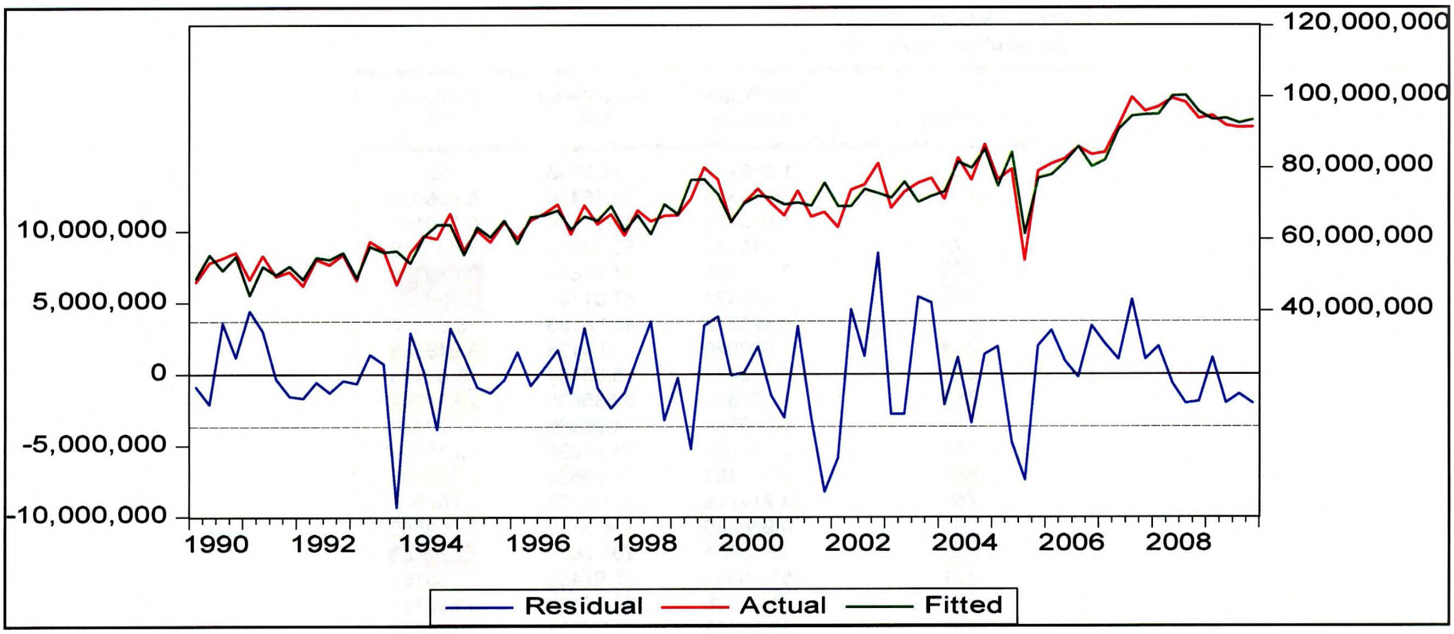


Table 10 – Equation 2: Ramsey RESET Test

Ramsey RESET Test:			
F-statistic	0.792978	Probability	0.503135
Log likelihood ratio	3.448923	Probability	0.327447

Test Equation:
Dependent Variable: CCTAX
Method: Least Squares
Sample: 1990:1 2009:4
Included observations: 80

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	3.01E+08	3.30E+08	0.912923	0.3653
X2	-89356694	1.12E+08	-0.795777	0.4296
X3	-17992948	22725186	-0.791762	0.4320
X4	-61402318	77003074	-0.797401	0.4287
Z11	-158.5311	201.3686	-0.787268	0.4346
Z21	339.0204	424.9165	0.797852	0.4285
Z22	-218.8979	273.6565	-0.799900	0.4273
Z23	-1.116453	1.515293	-0.736790	0.4644
Z31	0.005285	0.023612	0.223809	0.8238
Z44	-55.92273	69.93540	-0.799634	0.4274
Z48	-34.20066	43.07746	-0.793934	0.4307
Z51	-221.9832	279.6798	-0.793705	0.4308
Z52	-26.89904	33.80720	-0.795660	0.4297
Z53	2662.730	3346.902	0.795581	0.4298
Z54	-85.17855	106.1818	-0.802195	0.4260
Z55	144.0710	178.6278	0.806543	0.4235
Z56	2.689790	3.288708	0.817887	0.4170
Z61	-41.97618	52.87229	-0.793916	0.4307
Z62	-213.7105	267.1760	-0.799887	0.4273
Z71	162.6543	201.8559	0.805794	0.4239
Z72	-271.0010	339.9174	-0.797255	0.4288
Z81	-9.604231	13.75468	-0.698252	0.4880
Z92	27.15657	33.56064	0.809179	0.4220
FITTED^2	3.20E-07	4.11E-07	0.777508	0.4403
FITTED^3	-2.74E-15	3.87E-15	-0.706787	0.4827
FITTED^4	8.52E-24	1.34E-23	0.637251	0.5267

R-squared	0.952182	Mean dependent var	70221018
Adjusted R-squared	0.930044	S.D. dependent var	14063614
S.E. of regression	3719703.	Akaike info criterion	33.35314
Sum squared resid	7.47E+14	Schwarz criterion	34.12730
Log likelihood	-1308.126	F-statistic	43.01150
Durbin-Watson stat	2.011295	Prob(F-statistic)	0.000000

Table 11 – Equation 2: Serial Correlation Lagrange Multiplier Test

Breusch-Godfrey Serial Correlation LM Test:

F-statistic	0.116265	Probability	0.734398
Obs*R-squared	0.165748	Probability	0.683919

Test Equation:
Dependent Variable: RESID
Method: Least Squares
Presampple missing value lagged residuals set to zero.

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-12217.83	11142272	-0.001097	0.9991
X2	-7128.488	1573818.	-0.004529	0.9964
X3	13742.70	2197797.	0.006253	0.9950
X4	5406.419	1953128.	0.002768	0.9978
Z11	-0.211518	6.247347	-0.033857	0.9731
Z21	-0.039740	10.14669	-0.003917	0.9969
Z22	0.031471	3.965244	0.007937	0.9937
Z23	-0.013482	0.276623	-0.048738	0.9613
Z31	0.000555	0.021823	0.025433	0.9798
Z44	-0.091415	2.133746	-0.042843	0.9660
Z48	0.054197	0.980037	0.055301	0.9561
Z51	-0.422163	12.59852	-0.033509	0.9734
Z52	0.039280	1.416949	0.027722	0.9780
Z53	0.972441	23.68142	0.041063	0.9674
Z54	-0.249000	4.584695	-0.054311	0.9569
Z55	-0.392623	8.704102	-0.045108	0.9642
Z56	0.027897	1.675342	0.016652	0.9868
Z61	0.006737	1.518371	0.004437	0.9965
Z62	0.428148	5.684685	0.075316	0.9402
Z71	-1.685619	23.24538	-0.072514	0.9425
Z72	0.540171	8.236092	0.065586	0.9479
Z81	-0.079100	6.477352	-0.012212	0.9903
Z92	-0.234076	6.364759	-0.036777	0.9708
RESID(-1)	0.051491	0.151010	0.340976	0.7344

R-squared	0.002072	Mean dependent var	-1.35E-08
Adjusted R-squared	-0.407791	S.D. dependent var	3142340.
S.E. of regression	3728398.	Akaike info criterion	33.34418
Sum squared resid	7.78E+14	Schwarz criterion	34.05879
Log likelihood	-1309.767	F-statistic	0.005055
Durbin-Watson stat	1.977460	Prob(F-statistic)	1.000000

Figure 9 – Histogram for Equation 1: Regression Residuals

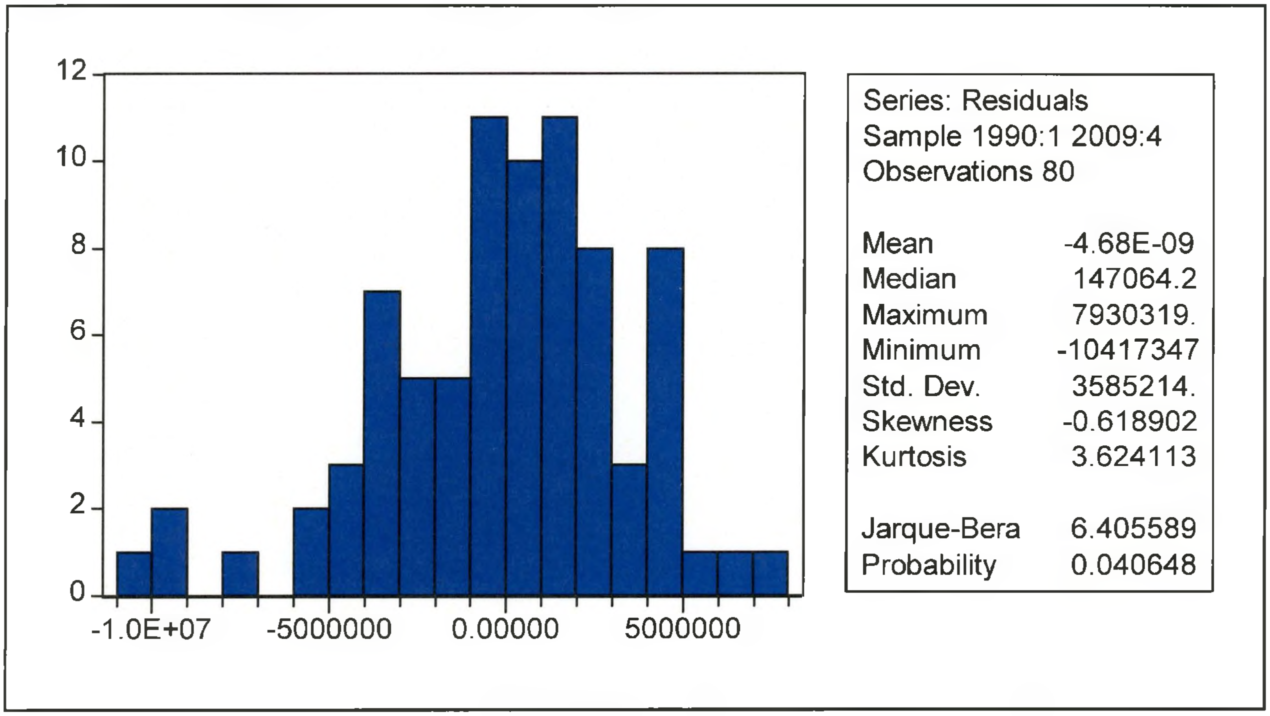


Figure 10 – Histogram for Equation 2: Regression Residuals