View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by DigitalCommons@USU

Utah State University

DigitalCommons@USU

Undergraduate Honors Capstone Projects

Honors Program

5-2015

Exploring the Relationship Between Utah's Wages and Utah's Real Estate Values

S. Scott Laneri
Utah State University

Follow this and additional works at: https://digitalcommons.usu.edu/honors

Part of the Finance Commons

Recommended Citation

Laneri, S. Scott, "Exploring the Relationship Between Utah's Wages and Utah's Real Estate Values" (2015). *Undergraduate Honors Capstone Projects*. 538.
https://digitalcommons.usu.edu/honors/538

This Thesis is brought to you for free and open access by the Honors Program at DigitalCommons@USU. It has been accepted for inclusion in Undergraduate Honors Capstone Projects by an authorized administrator of DigitalCommons@USU. For more information, please contact digitalcommons@usu.edu.

UtahStateUniversity
MERRILL-CAZIER LIBRARY

EXPLORING THE RELATIONSHIP BETWEEN UTAH'S WAGES AND UTAH'S REAL ESTATE VALUES

by

S. Scott Laneri

Thesis submitted in partial fulfillment
of the requirements for the degree

of

DEPARTMENTAL HONORS

in

Finance
in the Department of Economics and Finance

Approved:

Thesis/Project Advisor	**Departmental Honors Advisor**
Dr. Ryan Bosworth	Dr. Frank Caliendo

Director of Honors Program
Dr. Kristine Miller

UTAH STATE UNIVERSITY
Logan, UT

Spring 2015

Abstract

This paper uses a variety of multiple regression analysis techniques to attempt to answer whether a direct relationship exists between Utah's employee wages and Utah's residential real estate values. Unexpected declines in real estate values can have seriously negative impacts on businesses, individuals, and local governments in Utah. Conversely, unexpected increases represent missed opportunities. Researchers have used various statistical and mathematical methods to explain or predict changes in real estate values, but no method has consistently predicted values for a long period of time or across multiple geographical areas. This paper focuses on exploring the relationship between variables in Utah and uses a linear probability model with nine explanatory variables to attempt to explain trends in quarterly data from the Utah Housing Price Index over the last sixteen years.

Initially, the regression returned promising numbers, but the results were misleading. Due to nonstationary data, high levels of autocorrelation, and other issues related to time-series data, the regression results were spurious, and no useable conclusions were drawn from the first model. In an attempt to correct for autocorrelation and the nonstationarity, the variables were transformed using a Prais-Winston transformation. Again, the results appeared promising. After multiple tests for stationarity and autocorrelation, however, the results were found to be autocorrelated and spurious.

That being said, the time spent reading complex papers, gathering reliable data, researching advanced regression methods, transforming variables, re-specifying models and analyzing results has been a great help and will contribute to a solid statistical foundation in the future. Research opportunities are available in the future when higher level statistical methods are learned. A relation between Utah's wages and Utah's real estate values may exist, but the statistical methods necessary to create the proper model are beyond the scope of this paper.

Introduction

Exploring the relationship between wages and real estate values is important in Utah because entities and individuals in both the private and public sectors often earn (lose) money based on whether the value of real estate in Utah increases (decreases). In the public sector, for example, local governments rely heavily on revenue from property taxes to fund services. These services include public education, water sanitation, road repair, law enforcement, fire protection, utilities and a host of other services. Because property taxes are based on property values, an unexpected shift in the value of real estate can have a major impact on government revenues (Guan et al.,). To illustrate, consider the impact of an unexpected -2% change in Utah's primary residential real estate value. Primary residential real estate is taxed at an effective rate of .78% of fair market value. In 2013, Utah's total residential real estate value amounted to $167,285,244,324. This means that the total taxes levied from primary residential properties was $1,305,225,398. Now, apply the -2% change to the total value and calculate the new tax revenue. The total value of real estate drops by to over $3.3 billion to 163,939,539,437, and the tax revenue declines by nearly $26.5 million to $1,278,728,407 (Utah State Tax Commission 2014). Further consider that the reduced tax figure doesn't factor in the loss in value of other types of real estate, nor does it reflect a major swing in real estate values such as the downturns observed in the last decade. In 2009, for example, the Utah Housing Price index, which tracks the value of all residential home transactions in the state, moved -9.85%. Assuming this happened again, and using the same real estate values as in 2013, this would represent a decrease in state tax revenue by $130 million. This example shows that government tax revenues are highly susceptible to changes in real estate values. Therefore, a major decline in real estate values can create tough economic situations for local governments that may lead to a reduction in government services.

Aside from the public sector, shifting real estate values have major implications for the private sector. Companies and individuals often make investments because they speculate that the value of real estate will increase or decrease. Consider the following selections from a pro forma summary of a multi-family real estate development. The pro forma was created during an internship with the Ritchie Group, which is a real estate development and investments company based out of Salt Lake City.

Assumptions					
Market & Operating			**Project Financial Analysis**		
Vacancy		6.0%	Cost Yield-Stabilized-Yr 2		7.4%
Gross SF		289,800.00	Debt Yield Ratio-Stabilized- Yr 2		10.5%
Rentable SF		252,000.00	Debt Coverage Ratio-Stabilized-Yı		1.59
Average Resident $ / SF / mo	$	1.05	Total Expenses/Res Unit	$	4,044
Revenue Inflation		3.0%	NOI/Residential Unit	$	10,114
Expense Inflation		3.0%	Value Creation = (FMV - Cost Basis)		
Lease Up SF / mo		20,000.00	Yr 2-stabilized-Est FMV	$	41,428,800
Market Cap Rate		6.0%	Yr 6 Stabilized Est FMV	$	46,628,479
Ramp Up NOI	$	1,380,431	Yr 11 Stabilized Est FMV	$	54,055,187
2nd Yr Stabilized NOI	$	2,548,728	Year 2 Value Creation	$	6,866,770
3rd Yr Stabilized NOI	$	2,625,190	Year 6 Value Creation	$	12,066,449
Avg Rental Income / Unit / Yr	$	12,600	Year 11 Value Creation	$	19,493,157
Avg Other Income / Unit / Yr	$	2,314			
Avg Total Income / Res Unit / Y	$	14,158			
Avg OpEx $ / Unit / Yr	$	4,044			
Cap Reserves/Unit	$	250			

Projected Capital Structure								
			LTC		LTV	Rate		5.0%
Equity	$	10,368,609.08	30.00%			Term (in yrs)		30
Long Term Note A	$	24,193,421.19	70.00%		58.40%	Annual PMT		$1,558,506
	$	34,562,030.27				Monthly PMT		$129,876
			FMV Year 4	$	45,270,368			
Long Term Note B (Yr 4)	$	33,952,776.25			75%	Rate		6.0%
* Cash Distribution to Equity	$	8,756,652				Term (in yrs)		Coterminous w/ Note A
						Annual PMT	$	630,006.66
						Monthly PMT	$	52,500.56

Annual Cash Flow					
Year	1-Ramp Up	8	9	10	11
Potential Gross Income		$ 3,791,355	$ 3,905,095	$ 4,022,248	$ 4,142,916
-Vacancy		$ (227,481)	$ (234,306)	$ (241,335)	$ (248,575)
Collected Rental Income	$ 1,804,950	$ 3,563,874	$ 3,670,790	$ 3,780,913	$ 3,894,341
+Other Income	$ 331,481	$ 696,285	$ 717,174	$ 738,689	$ 760,850
Effective Gross Income	$ 2,136,431	$ 4,260,159	$ 4,387,964	$ 4,519,603	$ 4,655,191
-Total Operating Expenses	$ (756,000)	$ (1,216,844)	$ (1,253,350)	$ (1,290,950)	$ (1,329,679)
Net Operating Income	$ 1,380,431	$ 3,043,315	$ 3,134,614	$ 3,228,652	$ 3,325,512
-Reserve for Captial Imp.	$ (63,000)	$ (75,225)	$ (77,482)	$ (79,807)	$ (82,201)
Net Income	$ 1,317,431	$ 2,968,089	$ 3,057,132	$ 3,148,846	$ 3,243,311
-Debt PMT A note	$ -	$ (1,558,506)	$ (1,558,506)	$ (1,558,506)	$ (1,558,506)
-Debt PMT B note	$ -	$ (630,007)	$ (630,007)	$ (630,007)	$ (630,007)
Before Tax CFLO	$ 1,317,430.5000	$ 779,576	$ 868,619	$ 960,333	$ 1,054,798
+Proceeds from Refinance	$ -	$ -	$ -	$ -	$ -
+Net Residual Sale	$ -	$ -	$ -	$ -	$ 55,150,199
-Principal Payoff	$ -	$ -	$ -	$ -	$ (26,759,035)
Residual Before Tax Cash Flow	$ 1,317,431	$ 779,576	$ 868,619	$ 960,333	$ 29,445,963
Remaining Equity Invested	$ 10,368,609	$ 1,611,957	$ 1,611,957	$ 1,611,957	$ 1,611,957
Unlevered IRR 10.72%	$ 1,380,431	$ 3,043,315	$ 3,134,614	$ 3,228,652	$ 58,393,510
Levered IRR 22.25%	$ 1,317,431	$ 779,576	$ 868,619	$ 960,333	$ 29,445,963
Cash-On Cash Return	12.71%	48.36%	53.89%	59.58%	1826.72%
Yield on Cost	3.99%	8.81%	9.07%	9.34%	9.62%

Notice that there are over 25 assumptions in the pro forma. These assumptions affect every part of the transaction, from the loan amount and monthly payments to the expected IRR to the eleven years of cash flows. Also notice how the assumptions affect every line item in the eleven year cash flow model. For instance, it is assumed that rent, and subsequently revenue, will increase by 3% each year. This is important to note because commercial real estate is often valued by capitalization rates. To determine the value of a cash-flowing property, the annual net operating income (NOI) is divided by the capitalization rate. So, to determine the value of the property in year 11 from the documents above, the NOI of $3,325,512 is divided by the assumed capitalization rate of 6% and computes to a valuation of $54,055,183. But, assume that the 3% annual rent growth assumption is wrong. Instead, assume the revenue from this asset actually decreases by 2% each year. Then, in year 11, there exists an entirely different story. If revenues declined by 2% each year, the new net operating income in year eleven would be $1,644,980 and, assuming the same cap rate of 6%, the valuation would be $27,416,333, which is just over half the value using the original assumption. The change in revenue also moves the expected IRR from 24% to 4.4%. From this example it is obvious that small changes in real estate assumptions can have a major impact on a company's revenue. Thus, being able to find a correlation between trends in local wages and the value of local real estate in order to eliminate as much uncertainty as possible is extremely valuable to businesses, governments, and individuals.

Hypothesis

There is likely a positive relationship between increasing wages and increasing home values in Utah. As income increases, households increase their buying power and are able to devote more resources to buying a home and paying a mortgage. The expectation is that the data will reveal a positive contemporaneous correlation between Utah's Housing Price Index and Utah's wages.

Literature Review

Explaining and predicting the movement of real estate values is a topic that has received a fair, though not overwhelming, amount of attention in the academic world over the last twenty-five years. Since 1990, over 250 academic journal articles relating to predicting or explaining real estate values have been published in major journals. None, however, have been written specifically to predict or explain real estate values in Utah (Business Source Complete 2015).

Although this is an original topic for Utah, this paper is largely inspired from the methodologies and findings from the over 250 papers published on the subject. Researchers from around the world use a variety of methodologies and variables to predict or explain the changes in the value of real estate, but the two most common methods to predict or explain the movement in the value of real estate are multiple regression analysis (MRA) and artificial neural networks (ANN) (Nguyen and Al 2001).

Regression analysis is the most widely used method in predicting the changing value of real estate. MRA, in simple terms, is a statistical technique that is, "concerned with the study of the relationship between one variable called the explained, or dependent variable, and one or more other variables called the independent, or explanatory variables." (Gujarati and Porter 2010). Regression analysis uses preconceived rules or assumptions that are built into statistical models that are often used to predict or explain trends in large datasets. The accuracy of the predictions and explanations can vary based on the quality and the type of data, the variables, and the specification of the model. More specifically, researchers must address issues including nonlinearity, multicollinearity, and heteroskedasticity when using MRA (Nguyen and Al 2001). It is important for researchers to have an extensive working knowledge of statistics and the correct methods to use MRA because using incorrect techniques, gathering bad data, or interpreting the results in the wrong way can lead to highly misleading conclusions. When used correctly, MRA can be a powerful tool to predict or explain datasets relating to a wide variety of subjects.

The other common method used in predicting real estate values, the artificial neural networks method, is a complex computational method inspired from the way natural neurons process information. A simplified version of how natural neurons process information follows:

1) Natural neurons first receive signals through synapses attached to the dendrite or membrane of the neuron.
2) If the signals are strong enough, the neuron is then activated and will send a signal through its axon to another neuron.

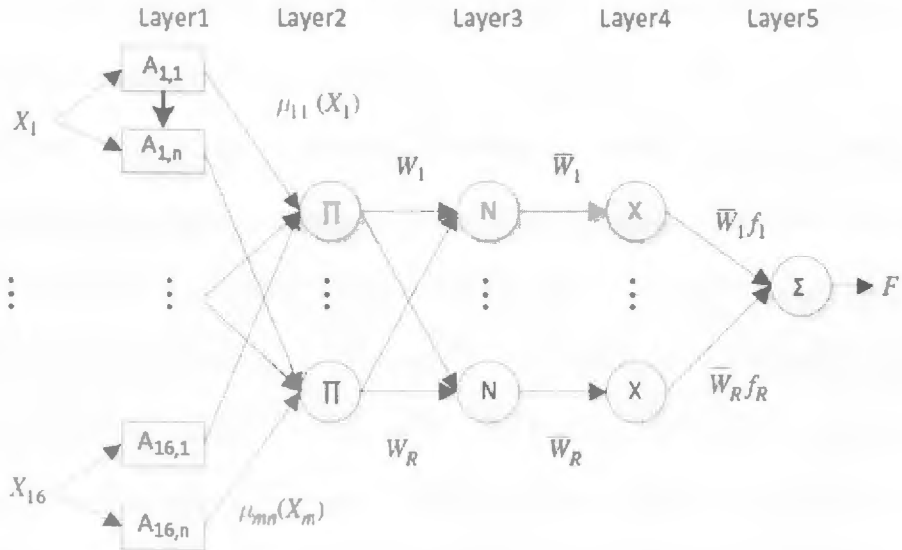
3) The combination of millions of neurons receiving and sending information is processed by the brain to make decisions on how to react.

Artificial neurons act in a similar way by using inputs, which are similar to synapses, and outputs, which are similar to axons. The strength of the signal received by the inputs of the neurons is determined and then multiplied by weights. Once multiplied by weights, a mathematical function determines the activation function that then determines the output of the neuron. All of this happens in milliseconds as algorithms and neural network software compute the data into interpretable output. Since the first neural model was introduced in 1943, hundreds of ANN models have been created. These models use hundreds or thousands of neurons to identify trends and patterns in datasets. ANNs can process large amounts of information to study animal behaviors, machine behavior, recognize patterns, and forecast data (Gershenson 2015).

Although ANN is an advanced method that is very useful in many applications, it has its setbacks and problems. Issues such as the number of hidden layers, number of neurons in each hidden layer, choosing of training set, the size of training set, selection of validation set, and overtraining all must be considered to create an effective model (Nguyen and Al 2001). As such, ANN should only be used by those who are experienced in using ANN methods because incorrectly using the method or incorrectly interpreting results will lead to misleading conclusions.

As shown, ANN and MRA are quite different in how they analyze data, but both methods are quite applicable in predicting real estate values. Mercedes A. Padilla, for instance, used MRA in 2005 to publish a paper titled *The Effects of Oil Prices and Other Economic Indicators on Housing Prices in Calgary, Canada*. In her paper, Padilla aimed to answer the following questions, "to what extent can oil prices and other economic indicators predict the changes in housing prices and rent in the Calgary single family housing market and [what is] the lag time between them." She used regression analysis to discover that up to 98% of the changes in house prices and rents can be explained by oil prices, exchange rates, interest rates, and employment levels. It was her paper that was the original source for the idea to explore if regression analysis based on Utah's local economic indicators could predict real estate values in Utah.

ANN also has successfully created models to predict real estate values. In a paper titled, *Analyzing Massive Data Sets: An Adaptive Fuzzy Neural Approach for Prediction, with a Real Estate Illustration*, Jian Guan, Donghui Shi, Jozef M. Zurada, and Alan S. Levitan introduce a variation of the artificial neuron network called the adaptive fuzzy neural approach to predict the value of homes for mass appraisals. The method uses 17 variables representing various attributes of a home to predict the value of the home. The variables include previous sale price, year built, year sold, square footage in the basement, number of fireplaces, garage size, number of baths, presence of central air, lot type, construction type, wall type, basement type, basement code, garage type, longitude, and latitude. The most striking difference between the variables in this model and most MRA models is that the ANN model uses actual coordinates of the home as a factor in the model. The coordinates allow the data to recognize the predicted value of homes on a particular street or neighborhood and to adjust values accordingly. The model used to produce the results is shown in the following figure:



Although the details of this model are complex, the intuition is quite simple. A piece of information, such as the type of garage or number of bathrooms, enters the model according to the variable it is related to and begins to pass through the layers of the model. Each layer contains neurons, also known as nodes, to receive and process information. The nodes process the input according to mathematical rules and assign each piece of information a weight or value. Once the value or weight is assigned, the node then sends the information to the next node. The new node goes through a similar process based on its own set of mathematical rules and then passes the information on once again. Once the information reaches the 5^{th} layer, the node computes the overall output (F) of the model as the sum of all the weighted outputs from the previous layers as shown below:

$$O = F = \sum_{k=1}^{R} \bar{W}_k f_k .$$

To more clearly explain the concept, consider the following example. A student named Kade is going to attend a new university in the fall, and he would like to register for a class as soon as he moves into his new apartment. Unfortunately, Kade has no idea what class to take.

Methodology

Data

The data used for this paper has been pulled almost exclusively from Federal Reserve Economic Research (FRED) datasets. FRED data is widely regarded as reliable and is commonly used in academic, government, and business studies. Although compiled by FRED, the data was reported quarterly by the U.S. Bureau of Economic Analysis, the U.S. Bureau of Labor Statistics, the U.S. Federal Housing Finance Agency, and other government institutions. Before transformations, the data was not seasonally adjusted and was measured in two units. The Housing Price Index uses the 1980 Q1 average home price as a base price of 100. The wage data was originally measured in thousands of dollars. As seen in Figures 1 and 2, nearly all of the raw data exhibits significant upward trends with shocks during the 2001 and 2007 recessions.

Figure1

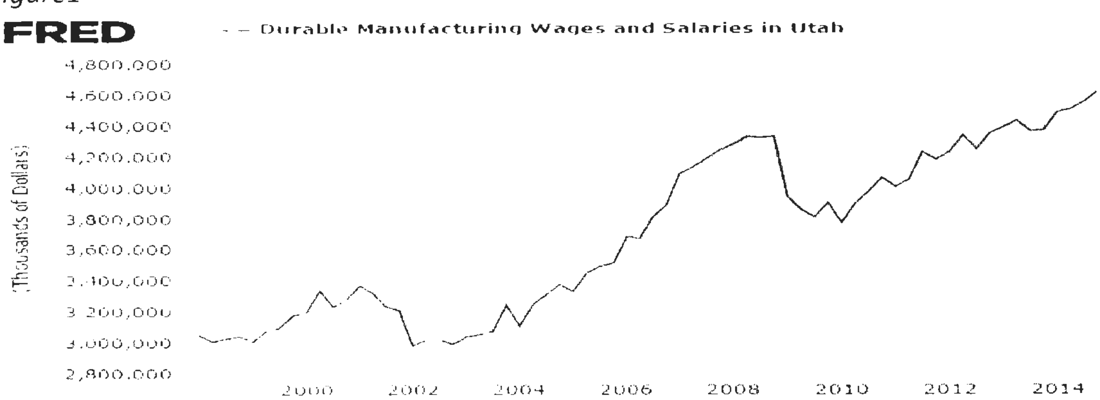


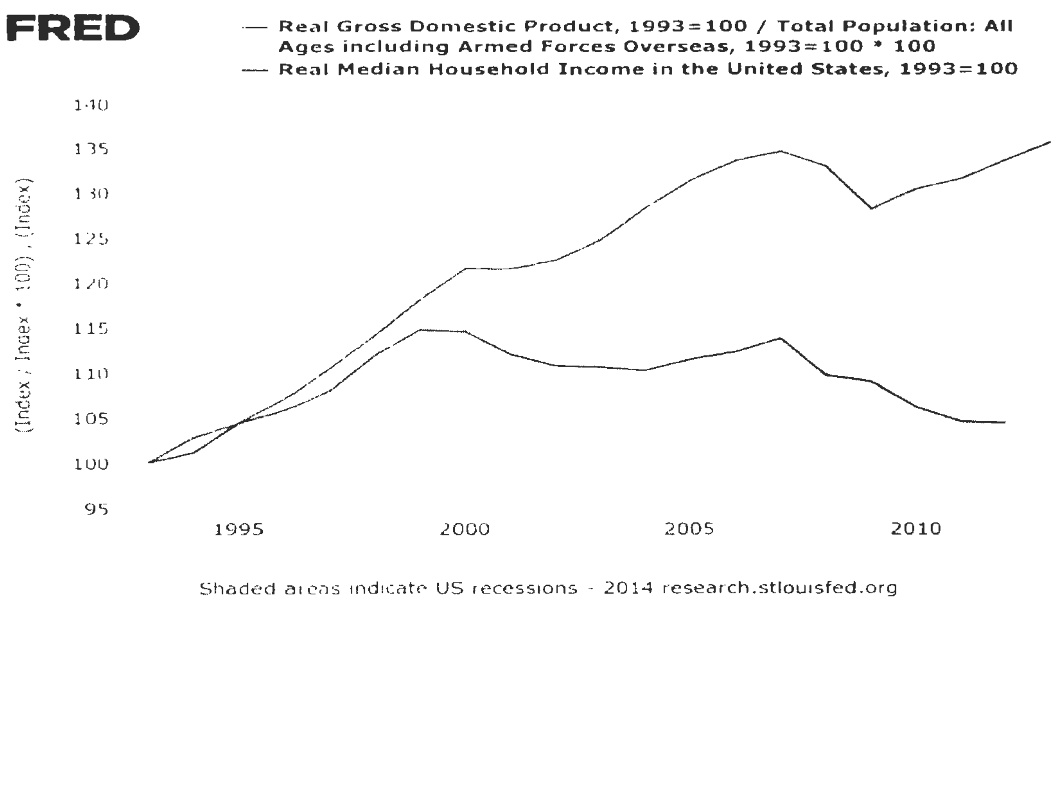
Figure2

Dependent Variable

The dependent variable in the model is the All Transactions Utah House Price Index. This data set contains 68 observations from 1980 Q1 to 2014 Q4. The Utah House Price Index was chosen as the measure of Utah's real estate value because it is one of the most consistent, accurate, and reliable sources available for Utah's real estate data.

Independent Variables

Although there are likely many variables one could use to predict real estate values in Utah, wages were chosen as the independent variable for this study because wages, more so than other macroeconomic indicators such as GDP or unemployment, would seem to better reflect the ability of Utah's residents to buy homes. Unemployment, for example, was specifically not chosen because it fails to measure if individuals are underemployed or if they have dropped out of the job search. Each condition would decrease home buying power. Consider also GDP; GDP was not chosen because it fails to accurately reflect how the overall growth in the economy is spread among individuals. In the United States, GDP per person doubled from 1970-2008, but the average income of American households as a percent of GDP shrunk during that same period. This is shown more clearly in the following graph:



The graph only illustrates the trend from the early 1990s, but it is evident that there is a growing disparity between GDP and the percentage of GDP that is reaching American households. Therefore, it shows that despite increasing GDP, the average American household is not benefitting as much as one might assume if only considering the economy's growth. GDP and unemployment can be good indicators of overall health and growth in an economy, but they fail to track how much of that growth the average person receives and the buying power of the individuals.

Wages track the buying power of individuals because wages represent dollars Utah workers are taking home each month. Although it is true that wages can potentially be deceiving if wages are inflated by a small segment of wealthy individuals, this, however, is why the explanatory variables represent different industries. The industries account for both high and low wage earners. Retail wages, for instance, will likely not be inflated by a small group of high-wage earners. On the other hand, the finance and insurance, management, and healthcare industries may inflate wages and not accurately reflect the average resident's home buying power. Even with inflated high-wage industries, the net effect of the high and low-wage-earning industries will provide a good representation of the population as a whole.

The independent variables in this model represent the quarterly wage totals from a variety of industries in Utah. The nine independent variables used in this model include:

1. Mining Wages
2. Accommodation and Food Services Wages
3. Durable Goods Manufacturing Wages
4. Finance and Insurance Wages
5. Education Services Wages
6. Retail Trade Wages
7. Health Care Wages
8. Management Wages
9. Other Services Wages

Figure 3 shows the nine industries' combined wages plotted against Utah's total wages and Figure 4 shows how the variables track the wage total in percentage terms.

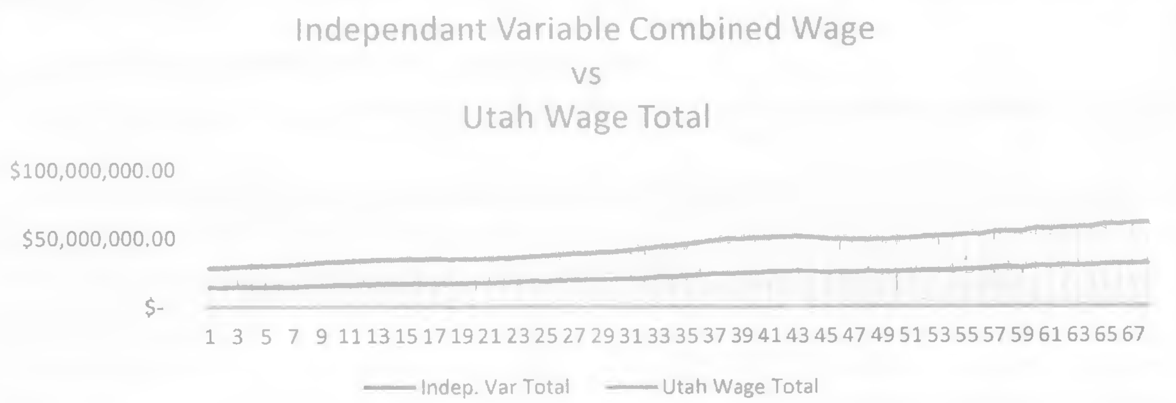


Figure 3

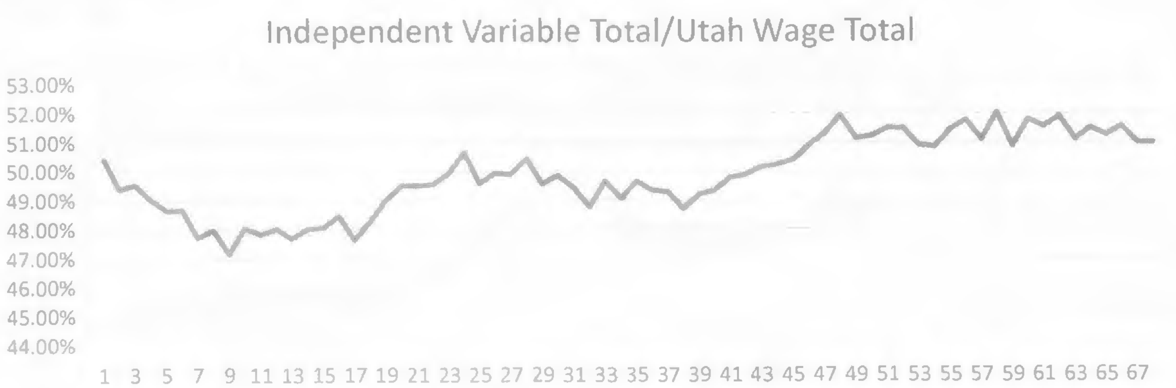


Figure 4

The variables were chosen because the combined totals represent approximately 50% of Utah's total wages. There are likely some industry variables that could be swapped out with another industry variable without changing the model significantly. But, the included industries were chosen because they were among the largest industries in Utah, and the industries create a good sample representation of Utah's low-wage and high-wage industries.

The first model uses a simple linear regression function:

$$Y=\beta+\beta_2 \text{MiningWg}+\beta_3 \text{AccmoFoodServicesWg}+\beta_4 \text{DurableManufacturingWg}+\beta_5 \text{FinInsuranceWg}+\beta_6 \text{EducationServicesWg}$$
$$+\beta_7 \text{RetailTradeWg}+\beta_8 \text{HealthCareWg}+\beta_9 \text{MangementWg}+\upsilon_i$$

Initial Results

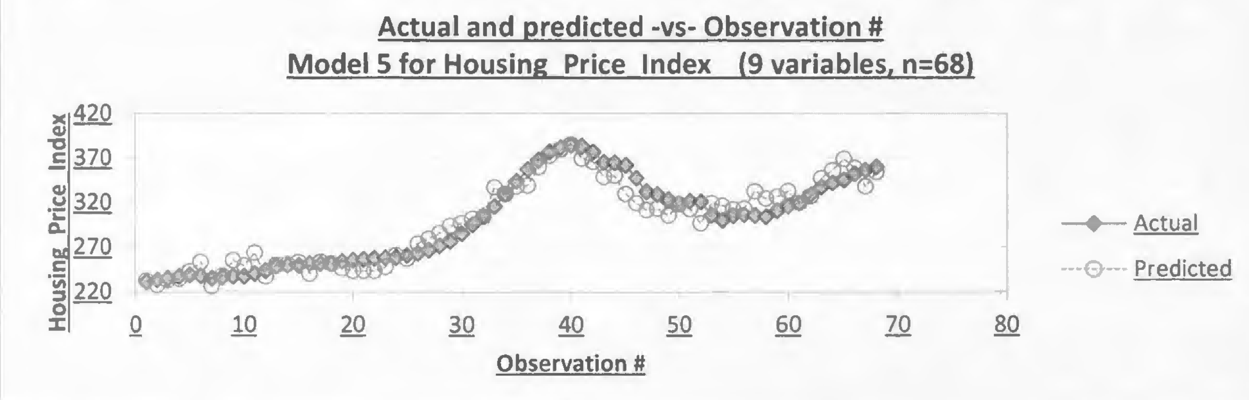
Regressing Utah's real estate values on the wages of some of Utah's major industries using the simple model returns the following results:

Regression Statistics: Housing Price Index (9 variables, n=68)

	R-Squared	Adj.R-Sqr.	Std.Err.Reg.	Std. Dev.	# Cases	# Missing	t(2.50%,58)	Conf. level
	0.924	0.912	14.384	48.468	68	0	2.002	95.0%

Coefficient Estimates: Model 5 for Housing Price Index (9 variables, n=68)

Variable	Coefficient	Std.Err.	t-Stat.	P-value	Lower95%	Upper95%	Std. Dev.	Std. Coeff.
Constant	126.793	52.924	2.396	0.020	20.855	232.732		
Accomo._and_Food_Services	0.000126	0.000107	1.168	0.248	-0.000090	0.000341	325,473	0.843
Durable_Manufacturing	-0.000032	0.000027	-1.193	0.238	-0.000085	0.000022	532,885	-0.349
Education_Services	-0.000050	0.000070	-0.711	0.480	-0.000191	0.000091	291,526	-0.301
Fin_and_Insurance	-0.000034	0.000025	-1.338	0.186	-0.000084	0.000017	679,267	-0.473
Health_Care	0.000014	0.000014	0.984	0.329	-0.000014	0.000042	1,109,057	0.319
Management_Wages	-0.000061	0.000022	-2.817	0.007	-0.000105	-0.000018	209,686	-0.265
Mining_Wages	0.000059	0.000044	1.338	0.186	-0.000029	0.000148	247,497	0.303
Other_Services	-0.000143	0.000039	-3.679	0.001	-0.000221	-0.000065	445,749	-1.318
Retail_Trade	0.000152	0.000029	5.303	0.000	0.000095	0.000209	650,342	2.039



Initial Analysis

At first glance, the results of the regression look very promising. The R^2 value is high, many of the t-stats are high, and the predicted values seem to track the actual values quite well.

As one looks deeper, however, some potential problems begin to surface. First, many of the coefficients are negative. This seems to go against the original assumption that an increase in wages would lead to an increase in the demand for homes, which subsequently would cause an increase in the price of homes. The assumption that increasing wages will increase real estate values is based on simple supply and demand economic theory: as demand for goods increase, the prices of those goods also increase. Perhaps a necessary variable has been omitted or a vital part of economic theory is not being grasped. Depending on whether or not the assumptions are correct, the signs of the coefficients that seem to go against economic theory create the first cautionary flag.

The second issue with the results is the high R^2 and high t-stats on some of the variables that are likely due to autocorrelation. Autocorrelation is common in economic time series data because the data has *intertia*. Inertia occurs because of business cycles. To illustrate, think of the economy as it begins to recover from a recession. The recovery process usually produces increasing values. Thus, the time series moves upward and the value of a series at one point in time is greater than its previous value. This means that time series data is likely interdependent or correlated to the previous value, which is generally a sign that autocorrelation will exist (Gujarati and Porter 2010).

Finally, the last issue with the results is that the Durbin-Watson d is lower than the R^2, and this suggests the results may be spurious. When working with time-series data, Granger and Watson have said that, "an $R^2 > d$ is a good rule of thumb to suspect that the estimated regression suffers from spurious (or nonsense) regression." So, although the R-Squared value is quite high and it appears that there is a definite relationship between wages in Utah and the value of Utah's real estate, a meaningful relationship may not exist. The issue with d and R^2 is likely occurring because the data used may be nonstationary data. As shown in figures 1 and 2, the data series are trending upward. This is generally a sign that the data may be nonstationary. In its weak form, data is considered stationary when, "its mean and variance are constant over time and the value of the covariance between two time periods depends only on the distance or lag between the two time periods and not on the actual time at which the covariance is computed," (Gujarati and Porter 2010). Unfortunately, if the data is not stationary, nonstationary data regressed on nonstationary data is often spurious.

Tests

Autocorrelation Test:

To test for autocorrelation, the Durbin-Watson *d* test is used. This statistic can be computed manually by analyzing the residuals in the model. The test is the ratio of the sum of squared differences in successive residuals to the RSS (Gujarati and Porter 2010). Fortunately, most software packages include the *d*-value in the regression results. The results are calculated with an excel add-on program called RegressIt (Nau, RegressIt Download 2015). The calculated *d* value from the first regression was .792. To test whether the calculated *d* value indicates that autocorrelation exists, the correct d_l and d_u must be found in the Durbin-Watson tables. The d_l and d_u for a model with nine explanatory variables and 70 observations are 1.337 and 1.910 respectively. Because the computed *d* value of .792 is far below the lower bound value of 1.337, there is positive autocorrelation between the explanatory variables and the Utah Housing Price Index. This means that the OLS estimators are no longer BLUE. In an effort to remedy the issue, the data will be transformed to obtain estimators that are generalized least squares estimators. The results with the transformed variables are found in the *Fine Tuned Model* section of the paper.

Unit Root Test:

To test whether the data is non-stationary, the unit root test will be conducted. This method is also known as the Augmented Dickey Fuller (ADF) test. To conduct the ADF test, the following regression is used.

$$\Delta y_t = \alpha + \beta y_{t-1} + \delta t + \zeta_1 \Delta y_{t-1} + \zeta_2 \Delta y_{t-2} + \cdots + \zeta_k \Delta y_{t-k} + \epsilon_t$$

In the ADF model, Δy_t is the first difference of Y and k is the number of lags used in the model (Stata 2015). To test the results of the first regression, the Augmented Dickey-Fuller unit root test is conducted on the dependent variable and one of the independent variables (all variables follow a similar looking trend) using an Excel add-in (Annen 2005).

The results of the ADF test for the dependent variable show that the t-Statistic is -1.5789 and the test critical values with 64 observations are as follows: 1%=-4.107983, 5%=-3.481605, and 10%=-3.168719. Because the calculated t-statistic of -1.5789 is far below the critical values, the null hypothesis cannot be rejected. Therefore, it is likely that the data contains a unit root, or in other words, the data is nonstationary. The ADF results for the independent variable Retail Trade show that the t-statistic is -2.7738. This critical value is also too small to reject the null hypothesis that the independent variable Retail Trade has a unit root. As with the dependent variable, the ADF results for the independent variable indicate that the data is nonstationary. This means that the regression was running nonstationary data on nonstationary data. Thus, the results of the first regression are very likely spurious.

Cointegration Test

Although the results of the first regression were found to be spurious, there may still be a long-run stable relationship between the dependent and independent variables. To test for cointegration, the residuals e_t from the initial regression must first be obtained. Then, treating e_t as a time series, another unit root test is conducted (Gujarati and Porter 2010).

The unit root test on the original regression's residuals return an Augmented Dicky Fuller test statistic of -2.274 and a p-value of .4639. Because the test statistic does not exceed any of the critical tau values, the data is not cointegrated and is nonstationary. This confirms that the data is spurious and that the results from the first regression are not useful for meaningful analysis or for forecasting.

Initial Key Findings

The results from the original model are disappointing. The output reveals little about the relationship between the explanatory variables and the dependent variable. The numbers initially looked promising, but problems with the data have created a massive hurdle that must be jumped in order to obtain results that are truly meaningful. The next section will address some of the issues in the model and attempt to correct them. If the errors can be corrected, meaningful results should be generated.

Fine Tuned Model

Using a simple model without any variable transformation doesn't reveal anything significant about the relationship between Utah wages and Utah housing values, but it is possible that after making some transformations to the variables, the results may be more meaningful. The lack of regression analysis expertise prevents running more complicated models to correct for errors. The model will, however, be adjusted to try and correct for autocorrelation using the Prais-Winsten transformation. This transformation is chosen because of the seriousness of autocorrelation. If the data can be corrected for autocorrelation, there is a possibility that meaningful results will be generated. The results of the regression are obtained from freeware software called gretl (Lucchetti and Cottrell 2015). In transforming the variables, the software estimates rho and then transforms the data according to the value of rho. The results of the regression are found below:

```
Model 2: Prais-Winsten, using observations 1998:1-2014:4 (T = 68)
Dependent variable: HousingPriceInd
rho = 0.992287

                      coefficient    std. error    t-ratio    p-value
  ---------------------------------------------------------------------
  MiningWages          1.14208e-05   2.23293e-05    0.5142    0.6091
  Accomo_andFoodS      3.11541e-05   4.05331e-05    0.7686    0.4452
  DurableManufact      8.65180e-06   1.01870e-05    0.8493    0.3991
  RetailTrade          2.55258e-05   1.38457e-05    1.844     0.0703
  FinandInsurance      1.75412e-05   1.16485e-05    1.506     0.1374
  EducationServic      3.44055e-05   2.65391e-05    1.296     0.1999
  HealthCare          -1.59850e-06   7.33590e-06   -0.2179    0.8283
  ManagementWages      1.48385e-07   8.75286e-06    0.01695   0.9865
  OtherServices       -1.28311e-05   2.05826e-05   -0.6234    0.5354

Statistics based on the rho-differenced data:

Mean dependent var   298.1556   S.D. dependent var   48.46845
Sum squared resid    2082.225   S.E. of regression   5.940703
R-squared            0.987019   Adjusted R-squared   0.985259
F(9, 59)             7.384386   P-value(F)           4.01e-07
rho                  0.373513   Durbin-Watson        1.189892
```

The new model increased R-squared from .924 to .987 and shows signs that the transformed data is slightly improved. The model also returned fewer negative coefficients, which helps connect the regression results with economic theory. Unfortunately, no individual explanatory variable is statistically significant and the Durbin-Watson statistic is still below the lower bound, which means that this data, although adjusted for autocorrelation, is still autocorrelated.

Fine Tuned Model Key Findings

Similar to the results from the first regression, the results from the second regression look fantastic at first but fail to maintain their statistical luster. This, unfortunately, means that not much can be drawn from these results because of the nature of the data used. If the results were valid, this would be a wonderful revelation to everyone investing in real estate. With a model that explains 98.7% of Utah's real estate values, companies, individuals, and governments would be much closer to understanding the movement of real estate values and would be able to make some very lucrative investments.

Conclusion

The time-series data used in this paper made it difficult to analyze using multiple regression techniques that were not beyond the scope of this paper. Although the results appeared promising on several occasions, the R^2 and t-stats were highly misleading. Thus, no groundbreaking results were found.

Although the results may not have provided much insight concerning the relationship between Utah's wages and Utah's real estate values, the process of researching methods to predict real estate values, gathering and manipulating data, creating a regression function, testing for problems in a regression model, fine tuning a regression model, and analyzing the results of a regression model have been extremely valuable.

In addition, learning what types of data are well suited for a specific type of analysis has proved useful. Time-series regressions are more suited for advanced, experienced statisticians than for an undergrad student in his first econometrics class. That being said, struggling with a time-series regression has its own merits. Although it is discouraging not finding meaningful results, time series regression is very interesting, and, with the right tools, it can be analyzed successfully.

Bibliography

Annen, Kurt. *[unit root test (adf-test) add-in].* 2005. http://www.web-reg.de/adf_addin.html.

Business Source Complete. *BSI Search.* April 30, 2015. http://web.b.ebscohost.com/bsi/search/advanced?sid=67c4f831-6b3f-4980-9a12-470c662599c8%40sessionmgr115&vid=26&hid=128 (accessed April 30, 2015).

Calgary Economic Development. *Economic Indicators.* 2015.

Calgary Meetings+Conventions. *Hold Your Event in Canada's Energy Capital.* April 16, 2015. http://www.meetingscalgary.com/www.meetingscalgary.com/business-sectors/energy.

Gershenson, Carlos. "Artificial Neural Networks for Beginners." *Datajobs.* 2015. https://datajobs.com/data-science-repo/Neural-Net-[Carlos-Gershenson].pdf (accessed April 30, 2015).

Gujarati, Damodar N., and Dawn C. Porter. *Essentials of Econometrics.* New York: McGraw-Hill Irwin, 2010.

Lucchetti, Riccardo "Jack", and Allin Cottrell. *gretl download.* 04 04, 2015. http://gretl.sourceforge.net/#dl.

Nau, Robert. *RegressIt Download.* 2015. http://regressit.com/download.html.

—. *Stationarity and differencing.* February 15, 2015. http://people.duke.edu/~rnau/411diff.htm.

Nguyen, Nghiep, and Cripps Al. "Predicting Housing Value: A Comparison of Multiple Regression Analysis and Artificial Neural Networks." *Journal of Real Estate Research*, 2001: 313-336.

Stata. *Augmented Dickey-Fuller unit-root test.* 2015. http://www.stata.com/manuals13/tsdfuller.pdf.

Utah GOED. *Targeted Industries.* 2015. http://business.utah.gov/industries/.

Utah State Tax Commission. *2013 Statistical Report.* Government Annual Report, Salt Lake City: Utah State Tax Commission, 2014.