

THE LIVING WAGE: TRENDS AND OBSERVATIONS

JAMES QUACINELLA

Capstone Project in Partial Fulfillment of the Requirements for the Masters of
Science Degree in Data Analytics (MS)

Submitted to the School of Professional Studies - City University of New York
June 2016

FOREWORD

This dissertation was done under the supervision of Professor Joshua Laurito at the School of Professional Studies in the faculty of Information Systems.

ABSTRACT

Amy Glasmeier proposed a model for the living wage [1], which she defines as:

The living wage model is an alternative measure of basic needs. It is a market-based approach that draws upon geographically specific expenditure data related to a family's likely minimum food, child care, health insurance, housing, transportation, and other basic necessities (e.g. clothing, personal care items, etc.) costs. The living wage draws on these cost elements and the rough effects of income and payroll taxes to determine the minimum employment earnings necessary to meet a family's basic needs while also maintaining self-sufficiency.

The original model produced data for 2014. The purpose of this project is to take this model and extend it's use to investigate trends in the living wage for the years 2004 - 2014. This project will look into what variables are most dominant, how the living wage distribution looks across the country, and how race and population affects the living wage distribution. Since the living wage is a measure of how much one needs to earn to meet basic needs, this project will also look at how many people are earning the living wage or less. The original model used 12 different family configurations. To keep the analysis simple, only households made up of a single person are modelled. Future work would expand this analysis to family configurations including children.

CONTENTS

1	INTRODUCTION	1
1.1	Introduction to the Living Wage and the Living Wage Model	1
2	DATA SOURCES AND COLLECTION	3
2.1	Consumer Expenditure Report	3
2.2	USDA Food Plans	3
2.3	Free Market Rent Data From HUD	3
2.4	Medical Expenditure Panel Survey from the AHRQ	4
2.5	Tax Data	4
2.5.1	State Tax Data	4
2.5.2	Federal Tax Data	4
3	MODEL VARIABLES	5
3.1	Housing Costs	5
3.2	Food Costs	5
3.3	Child Care Cost	5
3.4	Transportation Costs	5
3.5	Health Care Costs	5
3.6	Other Costs	5
3.7	Taxes	5
3.7.1	Payroll taxes	5
3.7.2	State Taxes	5
3.7.3	Federal Income Tax Rate	5
3.8	Final DataFrame	5
4	INTRODUCTORY ANALYSIS	7
4.1	Living Wage in Individual Counties	7
4.2	State Averages of the Living Wage	7
4.3	Choropleth Of Counties: 2004, 2006, 2014	7
4.4	Regional Averages of the Living Wage	7
4.5	National Average Breakdown Across 2004 - 2014	7
4.6	Living Wage Distribution in Most Populous Counties	7
4.7	Living Wage Distribution by Race	7
4.7.1	Loading Race Data from 2010 Census	7
4.7.2	Living Wage County (Non-Weighted) Breakdown by Race	7
4.7.3	Living Wage County (Weighted) Breakdown by Race	7
4.7.4	Population Weighted Averages Broken Down By Race	7
5	LIVING WAGE GAP	9
5.1	Data Collection	9
5.1.1	SAIPE - Small Area Income and Poverty Estimates	9
5.1.2	Minimum Wage Data	9

5.2	Distribution of The Median-Gap	9
5.3	Distribution of The Minimum-Wage-Gap	9
5.4	Percentage of Single Households At or Below the Living Wage	9
6	RESULTS AND FUTURE WORK	11
6.1	Overall Results	11
6.2	Future Work	11
I	Appendix	13
A	CODE LISTINGS	15
A.1	Data Sources	15
	BIBLIOGRAPHY	16

LIST OF FIGURES

LIST OF TABLES

LISTINGS

INTRODUCTION

1.1 INTRODUCTION TO THE LIVING WAGE AND THE LIVING WAGE MODEL

The living wage is defined by the original paper[1] as:

The living wage model is an alternative measure of basic needs. It is a market-based approach that draws upon geographically specific expenditure data related to a family's likely minimum food, child care, health insurance, housing, transportation, and other basic necessities (e.g. clothing, personal care items, etc.) costs. The living wage draws on these cost elements and the rough effects of income and payroll taxes to determine the minimum employment earnings necessary to meet a family's basic needs while also maintaining self-sufficiency.

The original model proposed estimated the living wage in terms of 9 variables:

$$\text{basic_needs_budget} = \text{food_cost} + \text{child_care_cost} + (\text{insurance_premiums} + \text{health_care_costs}) + \text{housing_cost} + \text{transportation_cost} + \text{other_necessities_cost}$$

$$\text{living_wage} = \text{basic_needs_budget} + (\text{basic_needs_budget} * \text{tax_rate})$$

The model, in summary, calculates the summation of common costs associated with basic living, and defines how much one needs to earn in wages to cover these costs (accounting for taxes). These variables have vary levels of coarseness: most variables are modelled at a regional level, with housing costs being the only variable modelled at the county level. This is a weakness of the model, and future work should focus on better per-county estimates of these variables.

We can also define a notion of the living wage gap, which is the difference between a household's income and the living wage. This gap is in some sense a measurement of how well a household can live above just above, or how much more income a household needs to meet, mere subsistence. This project will look at the difference between median wages earned and the living wage estimate at the county-level, and as well as the minimum wage.

The purpose of this project is to take this model and extend it's use to investigate trends in the living wage for the years 2004 - 2014. The project is structured as follows. The data sources that each of the model variables use are described in the Section 2, and the individual model variables are described in Section 3. Section 4 begins the analysis of the living wage distribution across the county. Section 5 looks at how we can compare median and minimum wage levels with the living wage to look at how well single households are doing with regards to meeting basic needs. Commentary will be made throughout the sections and summarized in the Results section, with extra commentary on where future work could go.

All code for this project can be found in the associated Github repository[2]. An alternative to this paper is the associated IPython notebook[3], which is also hosted in the Github repository.

DATA SOURCES AND COLLECTION

This section will outline how data was gathered for the various model parameters, as well as other data we need to calculate their values. The original model was made for 2014 data and extending this data to the past means we need to be careful that any changes in the underlying data methodology of these parameters needs to be noted. All data files mentioned here are available in the github repository, under the `data/` directory. Each data source is typically loaded into a Pandas DataFrame, which can be seen in the code sections linked to in the Appendix, or via the associated IPython notebook.

2.1 CONSUMER EXPENDITURE REPORT

The Consumer Expenditure Report[4] is used by the living wage model to determine 3 variables in the model, *transportation_cost*, *health_care_costs*, and *other_necessities_cost*. From their website:

The Consumer Expenditure Survey (CE) program consists of two surveys, the Quarterly Interview Survey and the Diary Survey, that provide information on the buying habits of American consumers, including data on their expenditures, income, and consumer unit (families and single consumers) characteristics ... The CE is important because it is the only Federal survey to provide information on the complete range of consumers' expenditures and incomes, as well as the characteristics of those consumers.

After downloading, the specific data needed for the model variables were extracted by hand. All data files are stored under the `cex_survey` subdirectory.

2.2 USDA FOOD PLANS

The Cost of Food project from the USDA[5] produces different food plans (The Thrifty, Low-Cost, Moderate-Cost, and Liberal Food Plans), which represent a nutritious diet at different costs. This dataset determines one variable in the model, *food_cost*. Also, the original model uses regional weighting factors to better model varying food prices across the county. [6] After downloading, the specific data needed for the model variables were extracted by hand. All data files are stored under the `food` subdirectory.

2.3 FREE MARKET RENT DATA FROM HUD

The U.S. Department of Housing and Urban Development produces the Fair Market Rent dataset, which the model uses as 'gross rent estimates' for the *housing_cost* variable. [7] After downloading, the specific data needed for the model variable were extracted by hand, using

the FMR1 column as the best estimate for the housing costs associated with a 1-bedroom apartment. All data files are stored under the housing_cost subdirectory.

2.4 MEDICAL EXPENDITURE PANEL SURVEY FROM THE AHRQ

What is MEPS, what is AHRQ, what its used for, link to code section in appendix

2.5 TAX DATA

2.5.1 *State Tax Data*

link to code section

2.5.2 *Federal Tax Data*

link to code section

MODEL VARIABLES

This section will describe each model variable, and link to a section of code in the appendix describing how its loaded.

3.1 HOUSING COSTS

Most Populous Counties

3.2 FOOD COSTS

What is USDA Food Plan, what its used for, link to code section in appendix

3.3 CHILD CARE COST

Currently, we are only looking at households that contain a single adult. Therefore, we do not model the costs of raising a child. One reason why this was done is that the data source for Child Care only goes back to 2006. Expanding on this work would find data for 2004 and 2005, and model the living wage for different family configurations.

3.4 TRANSPORTATION COSTS

3.5 HEALTH CARE COSTS

3.6 OTHER COSTS

3.7 TAXES

3.7.1 Payroll taxes

3.7.2 State Taxes

3.7.3 Federal Income Tax Rate

3.8 FINAL DATAFRAME

The final data frame, that includes each individual model variable as well as the total living wage *'total_{cost}'*, with a row for each county per year, is created in section bleh. This DataFrame is used by the following sections of analysis.

INTRODUCTORY ANALYSIS

This section will start to look at the living wage data to look for trends. First, we'll look into a few individual counties, and then look at state and regional averages. Secondly, we'll develop a set of maps of counties and their associated living wage. Finally, we'll look at the living wage distribution by population and by race.

4.1 LIVING WAGE IN INDIVIDUAL COUNTIES

4.2 STATE AVERAGES OF THE LIVING WAGE

4.3 CHOROPLETH OF COUNTIES: 2004, 2006, 2014

4.4 REGIONAL AVERAGES OF THE LIVING WAGE

4.5 NATIONAL AVERAGE BREAKDOWN ACROSS 2004 - 2014

4.6 LIVING WAGE DISTRIBUTION IN MOST POPULOUS COUNTIES

4.7 LIVING WAGE DISTRIBUTION BY RACE

4.7.1 *Loading Race Data from 2010 Census*

Move to data collection section

4.7.2 *Living Wage County (Non-Weighted) Breakdown by Race*

4.7.3 *Living Wage County (Weighted) Breakdown by Race*

4.7.4 *Population Weighted Averages Broken Down By Race*

LIVING WAGE GAP

This section will start to look at the living wage gap. First, we will come up with two definitions for the living wage gap, one based on the median wage, and one defined by the minimum wage. After that, we will look at distributions of both definitions. Finally, we will look at the distribution of households who earn the living wage or below.

5.1 DATA COLLECTION

Move this to data collection section

5.1.1 SAIPE - *Small Area Income and Poverty Estimates*

5.1.2 *Minimum Wage Data*

5.2 DISTRIBUTION OF THE MEDIAN-GAP

5.3 DISTRIBUTION OF THE MINIMUM-WAGE-GAP

5.4 PERCENTAGE OF SINGLE HOUSEHOLDS AT OR BELOW THE LIVING WAGE

Note that data is broken down into buckets, so numbers are an approx. Note the household definition again

RESULTS AND FUTURE WORK

6.1 OVERALL RESULTS

Will expand on each point:

- The living wage seems to have levelled off in last few years. Not sure why, but future analysis would be interesting
- Gap between minimum wage and the living wage reached a peak, and then some gains were made due to increases in the minimum wage. However, due to rent and inflation, current levels are close to the peak again.
- The top 150 most populous counties have a much higher living wage than the rest of the country. This is mostly due to rent
- The top 150 most populous counties can be split into two groups.
- White and blacks seem to have similar living wage distributions, as they are relatively well mixed across the county. Other races, especially asians and PI, are concentrated into areas with high living wages.
- Races seem to experience the same increases over time, with their distributions being controlled by population dynamics (since the living wage changes most with location)
- When looking at **only** single households, we see that on average, 50
- The region that has the most counties that have 50% or more of their single households making the living wage or less, is the South.
- Median wage?

6.2 FUTURE WORK

- If we could get wages broken down by race **and** by county, this would allow us to see how the living wage gap have evolved over time between races.

Part I

Appendix

CODE LISTINGS

A.1 DATA SOURCES

Code for loading all the data sources can be found in the IPython notebook under "Data Collection" [8]

REFERENCES

- [1] Eric Schultheis Amy K. Glasmeier, Carey Anne Nadeau. Living wage calculator: User's guide / technical notes. 2014.
- [2] James Quacinella. Associated github repository. 2015.
- [3] James Quacinella. Associated ipython notebook. 2015.
- [4] Consumer expenditure survey. 2015.
- [5] Usda food plans: Cost of food. 2015.
- [6] Stretching the food stamp dollar: regional price differences affect affordability of food. *Economic Information Bulletin Number 29-2*, 2007.
- [7] Fair market rents from the u.s. department of housing and urban development. 2015.
- [8] James Quacinella. Code for data collection. 2015.