Names: Christopher Chen & Sean Saito

ECON108 Regression Project: Milk Prices and Dairy Production

Our project aims to investigate the relationship between different dairy types, milk prices, milk production, and cow prices. The question we aim to answer is: "How do milk prices and cow prices affect milk production in the United States?". Based on the economic principles of supply and demand, we would assume that increases in milk prices would increase the production or supply of milk. This would possibly increase the prices of other dairy products as well. Also, the increases in prices of cows and the capital required to produce milk would also cause milk prices to increase.

In order to investigate milk production, we used a couple datasets that were compiled onto the TidyTuesday data project on GitHub. Those datasets were made up of data scraped from the USDA Economic Research Service which covers domestic supply and demand of various dairy products. Our data covers the dairy industry from 1980 to 2014. After merging fluid_milk_sales.csv and milkcow_facts.csv from TidyTuesday, there are 280 observations and 15 variables associated with each observation in our cleaned dataset. Each observation is a representation of one dairy type's production. The types of milk we have in the dataset are buttermilk, eggnog, flavored non-whole milk, flavored whole milk, low fat (%1), reduced fat (2%), skim milk, and whole milk.

For our analysis, we will be using simple linear regressions and fixed effect regressions. Two linear regressions will be performed: one with and one without controls. Here, milk production per year will be regressed on the average price of milk. Two fixed effects regressions will be performed: one with just entity fixed effects, based on milk type, and one with both entity and time fixed effects. Here we will be regressing the productions of different types of dairy products on milk prices. Our omitted milk type for the entity fixed effects would be buttermilk, and the omitted year would be 1980 for the time fixed effects.

In our first simple regression (1), the amount of milk produced in billions of pounds would be 78.604 if the price of milk was \$0, holding all else fixed. For every \$1 increase in milk price, the production of milk would increase by 574.167 billion pounds on average. This regression only had an R-squared value of 0.542, which is not very significant.

Our second linear regression (2) includes other variables as controls to better estimate the effect of price on milk production. This includes alfalfa hay price per ton, dairy ration price per pound, and cow slaughter price per pound of meat. In this model our constant is 117.023 billions of pounds of milk produced if the price of milk was \$0, holding all else fixed. Now, the price of milk is no longer statistically significant. Instead, alfalfa hay price had the most significant effect on milk price with a t-stat of 13.78. For every \$1 increase in hay price, milk production would increase by 0.950 billions of pounds. This regression had an R-squared value of 0.921, which is much more significant than the first model.

For our more complex regression (4), with both fixed effect variables, our constant would be 1,063.08 million pounds of product of buttermilk milk in 1980 holding all else constant. The coefficient of alfalfa hay price is 1.664 which means that when the price of alfalfa hay increases by \$1, the amount of buttermilk product increases by 1.664 million pounds. However, this coefficient is insignificant and this conclusion can't be stated. The average price of milk coefficient is -4,639.808, which means buttermilk product decreases by that amount when the milk price increases by $1, on average. The whole milk dummy variable coefficient shows that whole milk production is 19,616.866 million pounds greater than buttermilk. This coefficient is statistically significant with a t-stat of 20.81 and we can conclude that there is a lot more whole milk than buttermilk. For the time fixed effect, the dummy variable for 1981 is 1.199. This means that in the year of 1981, there was 1.199 more million pounds of buttermilk product holding all else constant. This coefficient is insignificant and this conclusion can't be stated. In model (3) without the time fixed effect, the alfalfa hay price and average price of milk slightly decrease. The constant slightly increases and the milk types stay the same but become more significant.

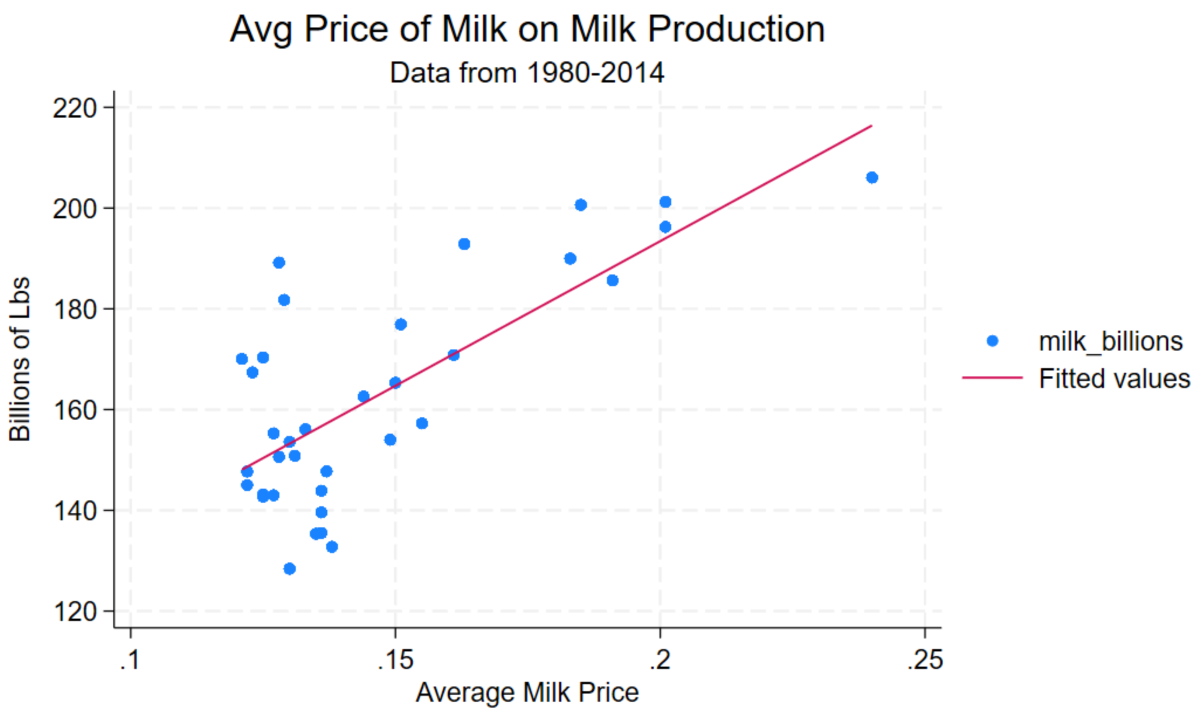
From our regression analysis, it seems like the most significant impacts on milk and dairy production are the prices of resources required to produce milk and the type of dairy product being produced. Between models (1) and (2), alfalfa hay price, price for dairy rations, and cow meat price had more significant effects on the production of milk than the price of milk itself. Looking at our time and milk type fixed effects regression, milk production or consumption patterns have not changed in the 35 year period since each year showed no statistically significant effects on amount produced per type. The conclusion is the same with the other variables alfalfa hay price and average price of milk with only milk type being statistically significant. This can be explained by consumers' preferences on milk. For example, someone who is lactose intolerant will always buy a certain type of milk regardless of an increase in price. Therefore, the supplier will keep the same ratios of product even if production costs increase.

Our regression models do suffer from omitted variable bias because we are unable to account for every variable that may affect milk production. For example, widespread cow diseases could affect cow milk production and milk prices. Cow disease is positively correlated with milk price and negatively correlated with milk production because the spreading of more diseases would cause milk production to slow and the price of milk to go up. Therefore, there would be a negative bias in our model because we do not take into account the health of the cows. The true milk price coefficient is biased away from zero.

Ideal variables that could have been used are data for each type of milk. For example, we only had the average price for all milks. Another variable is the amount produced for each state/region or country. Different areas have different diets and habits and dairy consumption patterns that could have been observed to see how their milk production is different. It would be difficult to create a model where X was randomized because our milk prices are determined by the market.

Names: Christopher Chen & Sean Saito

Figure 1: Average Price of Milk on Milk Production



Names: Christopher Chen & Sean Saito

Table 1: Milk Production Regression

VARIABLES	(1) Milk production (billions of lbs/yr)	(2) Milk production (billions of lbs/yr)	(3) Milk production (billions of lbs/yr)	(4) Milk production (billions of lbs/yr)
Alfalfa hay price ($ per ton)		0.950***	1.049	1.664
		(0.069)	(7.225)	(24.800)
Avg price milk ($ per lbs)	574.167***	2.504	-3,515.739	-4,639.808
	(24.464)	(41.613)	(8,607.366)	(30,357.154)
Dairy ration ($ per lbs)		-702.153***		
		(90.480)		
Slaughter cow price ($ per lbs of meat)		-27.779***		
		(6.278)		
Eggnog			-594.820***	-594.820***
			(34.062)	(54.328)
Flavored (Not Whole)			1,790.769***	1,790.769***
			(182.658)	(191.466)
Flavored (Whole)			40.311	40.311
			(38.695)	(58.582)
Low Fat (1%)			4,886.023***	4,886.023***
			(250.549)	(259.696)
Reduced Fat (2%)			16,929.871***	16,929.871***
			(321.636)	(330.354)
Skim			5,794.297***	5,794.297***
			(400.509)	(412.609)
Whole			19,616.866***	19,616.866***
			(872.705)	(942.815)
Constant	78.604***	117.023***	1,123.285	1,063.080
	(4.030)	(3.926)	(748.151)	(3,735.306)
Fixed Effects	None	None	Milk Type	Milk Type and Year
Observations	280	280	280	280
R-squared	0.542	0.804	0.921	0.921

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1