

# Report on Bachelor / Master Thesis

Institute of Economic Studies, Faculty of Social Sciences, Charles University

<b>Student:</b>	<b>Barbora Jakubová</b>
<b>Advisor:</b>	<b>Šarlota Smutná</b>
<b>Title of the thesis:</b>	<b>Determinants of demand for farm dairy products</b>

## **OVERALL ASSESSMENT** (provided in English, Czech, or Slovak):

*Please provide your assessment of each of the following four categories, summary and suggested questions for the discussion. The minimum length of the report is 300 words.*

### **Contribution**

This thesis presents an analysis of a market of farmers dairy products using as an wxample one specific firm from the Czech Republic. Detailed characterization of this market in the Czech Republic and demand analysis constitute a clear contribution to the literature. However, as written below, demand analysis would deserve some improvements.

### **Methods**

The empirical analysis presented in the paper consists of three parts. First, the author offers a detailed descriptive analysis of the market for dairy products in the Czech Republic and contrasts is with it with sales of the analyzed company offering farmers' products.

Second, the author estimates price elasticity of demand for dairy products sold by the analyzed company Mléko z farmy. This analysis is based on monthly time-series data for the period 2011-2018. Third, the author tries to analyze revealed consumer preferences using data on individual consumers of Nákup z farmy collected in March 2020.

Descriptive analysis is quite well presented. Several interesting observations are made. On the other hand the author sometimes tends to overinterpret. For example, she explains the decreasing thrend in sales of Mléko z farmy (Figure 3.3) as: "The demand for farm products was also reflected by supermarkets or eshops, who also extended their offer. Hence, increasing competition influences the decline in sales as well."

Estimation of price elasticity of demand is performed by a simple OLS where the dependent variable is quantity of a given product sold in a given month by Mléko z farmy and the explanatory variable is proce of this product (and its two lags). The author also adds several control variables to deal with seasonality and to productt availability. She also controls for the average per capita consumption of the given product and households expentitures on food in CR. Model assumptions are clearly listed and most of them are discusses. The author runs tests for stationarity, multicollinearity, autocorelation, and heteroskedasticity and takes proper action according to these tests results. She however completly ignores the zero conditional mean assumption, which is often not satisfied when estimating demand equations, because price is not exogenously set.

To analyze consumer preferences the author designs a questionnaire (at least I guess she has designed it, is is actually not clearly written in the thesis who designed it and how consumer e-mai laddresses were collected so that the questionnaire could be sent) to collect the necessary information. The questionnaire was distributed in the second half of March 2020 and was answered by 468 individuals. Unfortunately, it could have been influenced by the outbreak of the pandemic, but the author could do nothing about is (well, she could at least comment on it). According to Figure 3.5 in March 2020 there were about 4000 returning consumers, so the questionnaire was filled in by just a tiny fraction of them. How does the author deal with selection? Is there any comparison of the sample of consumers who filled in the questionnaire to the sample of all consumers? For example – are loyal consumers oversampled? Or any other type of consumers - e.g. those supporting local food? The composition of the sample can largely affect results of the analysis, and yet it is not discussed in the thesis. Second, as I write in more detail below, it would be good if the design of the questionnaire (e.g. formulation of questions) was inspired by the model that will be estimated, no the other way round (the author writes: „After analysing the responses from the questionnaire, some answers are not

# Report on Bachelor / Master Thesis

Institute of Economic Studies, Faculty of Social Sciences, Charles University

<b>Student:</b>	<b>Barbora Jakubová</b>
<b>Advisor:</b>	<b>Šarlota Smutná</b>
<b>Title of the thesis:</b>	<b>Determinants of demand for farm dairy products</b>

convenient to be employed as independent variables in the models, for example, some factors - quality, origin, freshness.“

Data from the questionnaire were matched with purchases data from the firm's system. These data cover the period up till April 2020, what included the huge spike in food e-shops activities due to the lockdown. Given the scope of the thesis, it would be much better to exclude this period from the analysis – i.e. not count the purchases made after March 12 and not count consumers who have newly registered after this date. Behavior of consumers during this period is uncomparable to their earlier behavior. Of course, it would be interesting per se to see who and how changed their buying behavior during lockdown.

Using these data the author estimates two types of models. First, she tries to determine which consumer characteristics predict activity in the e-shop Nákup z farmy. As dependent variables she uses dummies representing frequency of shopping. Whether a person placed any order (1st model), whether a person placed more than one order (2nd model), whether a person placed 10 or more orders (3rd model). Probit model is used to estimate these models, what is a correct choice given that the dependent variables are binary (although the argument „probit model is used by economists and political scientists because it can deal with the problem of heteroscedasticity“ is not correct in the context of comparing probit with logit). Given that binary response model is used, the equation 4.7 is not correct, as it represents a linear model. The choice of explanatory variables is not well supported and it seems that the author did not have any microeconomic model in mind when writing down these equations. Here I observe a mix of variables that represent consumer characteristics that have a potential to affect their buying behavior and variables that actually reveal their buying preferences. This makes it difficult to interpret model estimates. Moreover, as written above, the dataset is probably very selected, only very specific customers filled in the questionnaire. Consequently, model estimates do not give much information.

Finally, the author tries to predict average spendings per purchase and the probability to include specific products in their baskets. These models, unfortunately, suffer from similar problems as described above. Namely sample selection and uninformed choice of explanatory variables. For example, the author tests the following hypothesis: „Those who are used to frequent buying on the internet are more likely to place a higher number of orders“ by including an explanatory variable binary variable equal to 1 if a person buys foodstuff very often on the internet. This could however lead to spurious correlation, as placing many orders on Nákup z farmy actually means that one is often buying food on internet.

Similarly like in the descriptive analysis, also when interpreting regression results the author tends to overinterpret. For example: „It seems that there is no difference between most of the products purchased by retired people; only eggs are 14% less likely to be bought by them. The tendency of older people to keep hens is still high, or they may have own source (friends, neighbours) of home eggs and do not need to buy them.“

## Literature

The literature review section is, together with the introduction, the best chapter of the thesis. Despite some of the issues described below, I consider the literature review as informative and relevant for the presented analysis. Also, it should be appreciated that the author compares her results to results of other studies in further chapters.

## Manuscript form

The manuscript has a very pleasant graphical form and is written in good English. On the other hand, it would benefit from reorganization, as information is sometimes provided in a bit chaotic way, especially in the second half of the thesis.

# Report on Bachelor / Master Thesis

Institute of Economic Studies, Faculty of Social Sciences, Charles University

<b>Student:</b>	<b>Barbora Jakubová</b>
<b>Advisor:</b>	<b>Šarlota Smutná</b>
<b>Title of the thesis:</b>	<b>Determinants of demand for farm dairy products</b>

For example, the section 2.1 is a mixture of discussion of a theoretical framework used to motivate the empirical analysis and a literature review of related theoretical concepts. It would be very useful for the reader and for the student herself to devote a separate chapter to discuss the theoretical/analytical framework where she would clearly state the goals of her analysis and relate them to economic theory of demand (and supply!!). This would greatly help with designing the questionnaire used to collect consumer data and with estimations.

Further, the methodology and data sections are somehow disorganized. Although most of the relevant information is provided, there is no clear flow of thoughts. I would suggest to first define what and generally how will be estimated and then discuss what data is necessary to populate the regression models. This would nicely motivate questions included in the questionnaire. Currently it is not clear how the questionnaire was designed.

Finally, there are some minor inconsistencies in the text, which however do not affect the general reading of the thesis.

## Summary and suggested questions for the discussion during the defense

This is an interesting bachelor thesis and despite the weaknesses described above, it fulfills the requirements for a bachelor thesis at IES, Faculty of Social Sciences, Charles University, I recommend it for the defense and suggest a grade C.

The results of the Urkund analysis do not indicate significant text similarity with other available sources.

I suggest discussing the following during the defence:

1. Why might the price of milk be endogenous in the demand model? What are the consequences of this in terms of estimation results? What can you do to overcome this problem?
2. Can you compare the sample used in probit regressions with the whole population of Nákup z farmy consumers? Does it differ? If yes, then which characteristics are different? How does it affect the conclusions based on your analysis?

## SUMMARY OF POINTS AWARDED (for details, see below):

CATEGORY	POINTS
<i>Contribution</i> (max. 30 points)	25
<i>Methods</i> (max. 30 points)	15
<i>Literature</i> (max. 20 points)	20
<i>Manuscript Form</i> (max. 20 points)	15
<b>TOTAL POINTS</b> (max. 100 points)	<b>75</b>
<b>GRADE</b> (A – B – C – D – E – F)	<b>C</b>

**NAME OF THE REFEREE: Barbara Pertold-Gebicka**

**DATE OF EVALUATION: 27.8.2020**

Digitálně podepsáno (27.8.2020)  
Barbara Pertold-Gebicka

---

**Referee Signature**

**EXPLANATION OF CATEGORIES AND SCALE:**

**CONTRIBUTION:** *The author presents original ideas on the topic demonstrating critical thinking and ability to draw conclusions based on the knowledge of relevant theory and empirics. There is a distinct value added of the thesis.*

**METHODS:** *The tools used are relevant to the research question being investigated, and adequate to the author's level of studies. The thesis topic is comprehensively analyzed.*

**LITERATURE REVIEW:** *The thesis demonstrates author's full understanding and command of recent literature. The author quotes relevant literature in a proper way.*

**MANUSCRIPT FORM:** *The thesis is well structured. The student uses appropriate language and style, including academic format for graphs and tables. The text effectively refers to graphs and tables and disposes with a complete bibliography.*

**Overall grading:**

TOTAL	GRADE
91 – 100	A
81 - 90	B
71 - 80	C
61 – 70	D
51 – 60	E
0 – 50	F