Preparing a dissertation proposal: Writing the Introduction, Problem Statement, Purpose Research Questions, and Hypotheses

HAYDAR KURBAN

PROFESSOR

DEPARTMENT OF ECONOMICS

HOWARD UNIVERSITY

HKURBAN@HOWARD.EDU

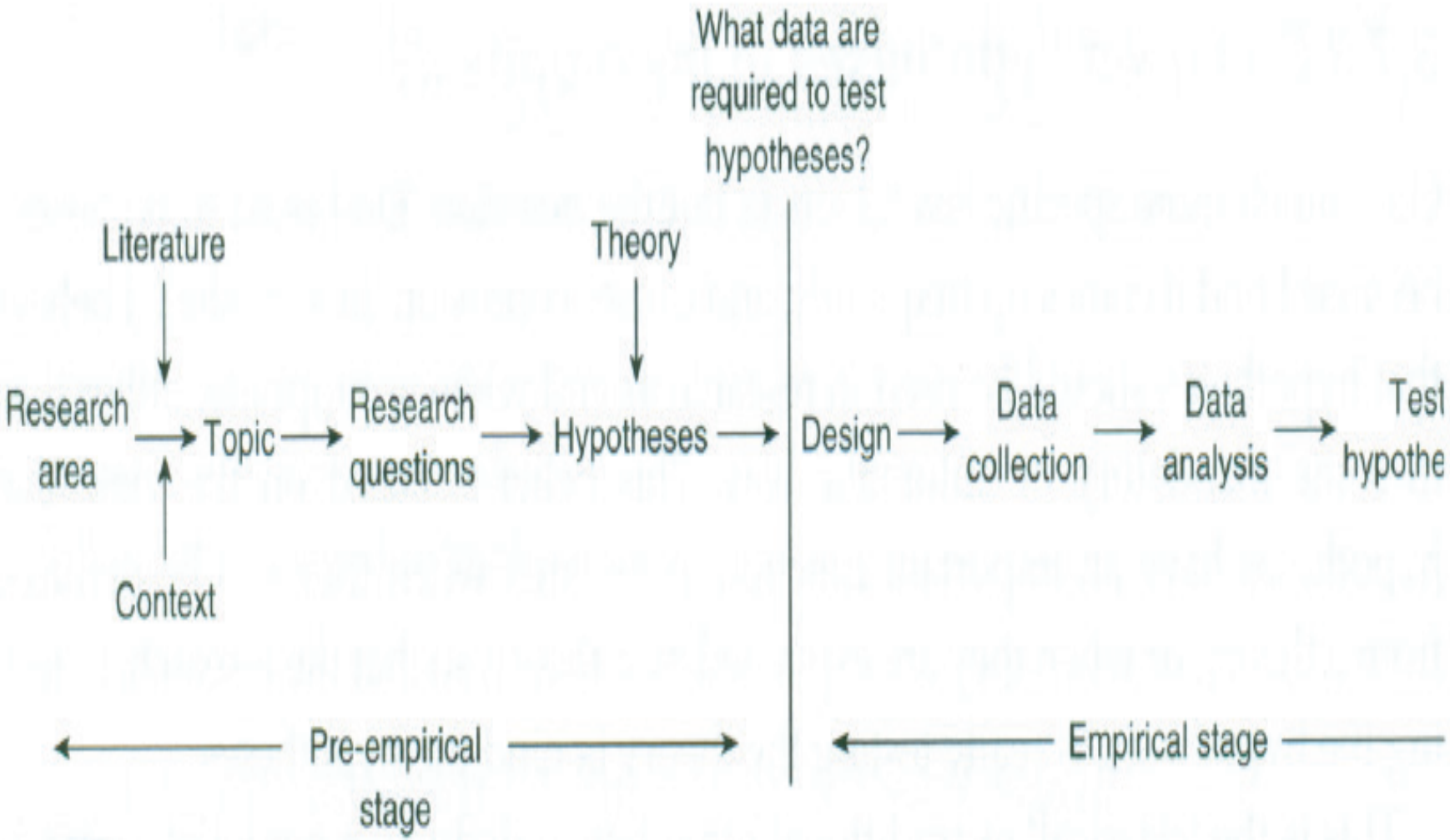


Figure 3.1 *Simplified model of research (with hypotheses)*

Punch, K. F. (2007). *Developing Effective Research Proposals* (2 ed.). Los Angeles, CA: SAGE Publications.

Essentials of Research Proposal

- An outline of the research problem
- Background of the problem (motivation)
- A statement of the significance of the problem
- Review of related studies (literature review)
- Methods for data collection, methods and analysis
- Statement about the study's contributions
- Statement about the study's limitations

Potential Sources of Research Proposal

- Reviewing published and unpublished research
- Questioning previous research results and developing a new study to test its findings
- Finding gaps or missing links in existing literature, theories and methods
- Designing and developing new research tools, models and techniques
- Implementing a known or tested method to new areas
- Trying to solve "real world" problems
- Interacting with with researchers in other fields (interdisciplinary research) to share ideas and methods

Review of Related Literature – Important Information

- Provides familiarity with the field of study
- Provides a context and background for the proposed study
- Generates research questions or hypotheses for the proposed study
- Shows the shortcomings of the existing methods, data and theories
- May lead to discovering new data sources or methods to address these shortcomings
- Provide knowledge of the methodologies commonly used in the study area
- Can reveal sources of data that were not known
- Enable you to demonstrate the timeliness and importance of the proposed study
- Learn from others' experiences to avoid or overcome roadblocks

Research Questions

- How to construct and test the validity and scope of research question before the work begins
- Research question cannot be too general
- Make sure that data and methods exist to investigate the question
- Modify or change research question based of feedback from literature review
- Derive testable hypotheses from research question
- Research question should be interesting (statistical significance versus relevance)
- Provide theoretical background or support for research question

Writing the Methodology

- how to select a research methodology

1. Review of relevant theories and literature is key in selection of research methodology
2. Implementation of a novel method itself can be the main contribution of the research

- how to use empirical data to substantiate ideas

1. Review of literature reveals the weaknesses and challenges of existing data sets
2. A data sets with more data points or better coverage (geographic or demographic detail) can be an improvement

- how to write the methodology section

1. A well defined research question supported by existing theories and knowledge (literature) is a prerequisite
2. Establish link between research question and appropriate method
3. Clearly write the steps involved in implementation of the method

- how to substantiate ideas with empirical work

Writing Methodology, Empirical

- How data set was created?

1. What data sources were used?
2. Why these data sources were appropriate?
3. Establish links to the literature: advantages or novelties of your data
4. How did you choose your variables from your data set?
5. How did you create your indicators and variables from the raw data?
6. What types of statistical validity tests you used in your analysis?
7. What types of robustness checks you used?

Writing Methodology, Analysis

- How did you analyze your data to answer your research questions?

1. What empirical method or methods you used?
2. Explain why your method is appropriate and how it fits into the literature.
3. Your empirical method may require modification of your variables
4. Explain why you used binary variables, logarithm forms and others
5. Make sure that you write down all your assumptions
6. Note that regression methods are compared in terms of the rigidity of the assumptions of each model

LIN LIN, LIN LOG, LOG LIN, and LOG LOG

- LOG-LOG Form

$$\log(Y) = \beta_0 + \beta_1 \cdot \log(X) + \varepsilon$$

- What do our coefficients say?
- The estimated value of β_1 tells us the effect of a 1-unit change in $\log(X)$ on $\log(Y)$.
- If X changes by a unit percent percent, Y changes with β_1 unit percent
- The relationship between X and Y can be explained in terms of elasticities

LOG-LIN,LIN- LIN, and LIN-LOG

LOG-LIN form

Log(Y) = $\beta_0 + \beta_1 X + \varepsilon$

- A one unit change in X will change Y by β_1 percentage

LIN-LIN form

- Y = $\beta_0 + \beta_1 X + \varepsilon$
- A 1 unit change in X will change Y by β_1 unit
- LIN-LOG form

Y = $\beta_0 + \beta_1 \text{Log}(X) + \varepsilon$

- A 1 percent change in X will change Y by β_1 unit

Data Sources and Retrieval

- identifying data sources

1. Since a new data source itself can be a contribution of a research project, identifying data sources is also part of literature review
2. Data sources, problem statement and methods can be jointly determined

- study data

1. Start to explore codebooks or or other tools to understand how well the data describe the study area
2. Confidentiality issues may limit geographic coverage (For example in Census micro samples a person's block group or census tract can not be identified)
3. Some variables may be top coded

Example: Payday Lending Study

- ***How did we develop the research proposal?***

- Why so many states have passed payday – friendly legislation?

1. Start to read papers, reports and regulation history
2. What do studies show?

- Many studies document that payday loans harm consumers?
- Both industry advocates and consumer advocates provide economics arguments for and against the economic benefits of the industry
- Data sources: CPS supplement 2009, 2010, 2011, 2013 and 2015
- PEW research reports and CFPB reports indicate that there is demand for payday loans
- Consumers find them very convenient compared to bank loans
- Reformulate research question

Reformulate Research Question

- Literature review indicates that payday customers have income in the range of 15K to 75K
- People cross state borders when one state bans payday loans
- Most of the payday customers have bank account (because you need one to get payday loans)
- Payday borrowers have impaired credit history
- Payday lenders charge the maximum allowable fees in every state
- Competition does not work or does not bring down fees
- On average a payday borrower takes payday loans 10 times in a year

Modify Research Question

- What is the net economic benefit of payday lending on a state economy?
- There are many studies on whether consumers benefit or being harmed?
- Many arguments made by the industry are not supported by the literature
- Payday industry harms lower income population but it may have net negative economic impacts on a state economy
- There Is one industry report that overestimates the net impacts but it is methodologically flawed
- I use economic impact models (IMPLAN) developed for 4 states

Now Clear on Research Question, Data and Method

- NAICS classification of payday lending is too general
- My model over estimate the positive impacts
- In a way it is good because it establishes an upper bound
- I estimate my model with IMPLAN
- I started with a naïve research question and based on literature review I modified it
- The new research question goes beyond the existing literature
- It adds another dimension
- The industry's impact on borrowers is well documented

Looking at another Aspect of Payday Lending

- Some studies mentioned the negative impacts of payday lending on state economies
- Payday borrowers experience higher degree of financial difficulty and higher default rates
- Since payday borrowers are very small portion of a state population, these studies only indirectly brought out some direct negative impacts
- My paper tries to estimate direct, indirect and induced effects on state economies
- I identified two channels: payday industry generate some positive economic impacts through directly and indirectly creating new employment

Another Insight from The Literature

- On the negative side industry charges more fees than the actual loan amount
- The industry's profit maximizing strategy is based on lending the same person many times
- On average a borrower pays $600 fees on a $300 loan.
- To pay these fees consumers cut their spending on health, education and food (well documented by the literature)
- Fees support profits and profits drain local consumption
- Our results show that the negative impacts dominates the positive ones
- Unanswered question (future work): why a state supports an industry that not only hurt the lower income population but also has overall negative impacts on the state economy

Public Data Sources

- Limitations of public data

- Privacy issues limit data availability at finer geography (example: Census tract or block groups)

- Privacy issues limit availability of some variables (top coding, grouping and missing observations)

- Remedies: Use GIS to combine data from different geographic details (example: concentration of crime incidences, fast food places around neighborhoods)

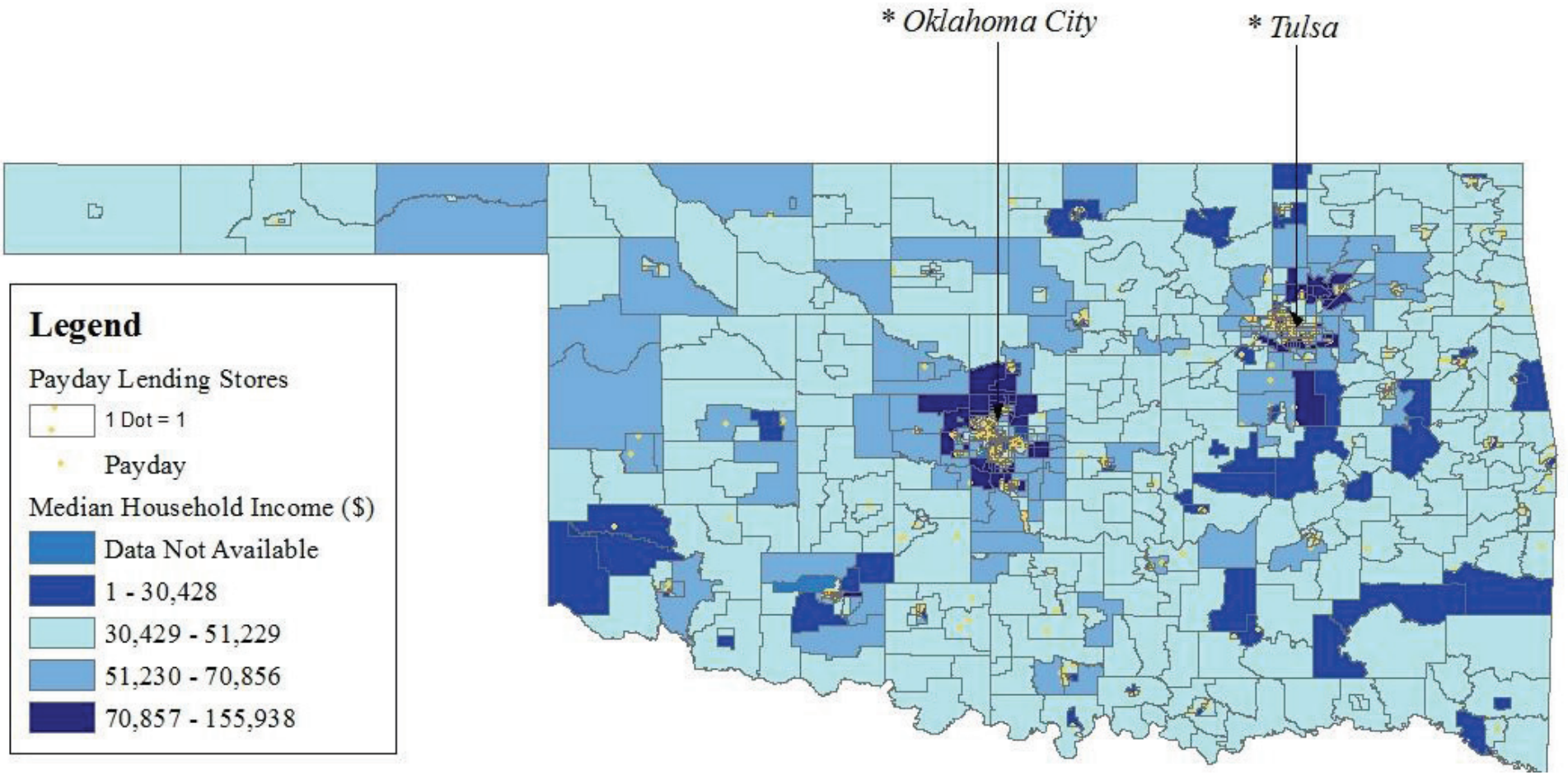
- You need to learn how to use Data Visualization tools (GIS and other spatial data sources and tools

-) Creating and using Synthetic Data Sets (example, we created Census block group level Micro Samples by using heuristic methods such as hill climbing and proportional fitting procedure in Kurban et al 2012.

Going beyond public use data and administrative data

- Data scraping (extracting data from websites)
- Many research papers and new dissertations scrap data from various websites (example: what type of restaurants survive in cities? Scrap menu and demand from restaurant websites)
- Scrap data from google search, facebook and twitter (example: assessing public sentiments during an event such as disasters, elections, or big demonstrations)
- Big Data tools: R and beyond (Example: We extracted 3-day and 7-day local weather forecast data from National Weather Service by using R)
- Increasingly Census Bureau and other data sets are supplemented by R codes. One can create variables and perform analysis by using a comprehensive R script.

Figure 1: Dispersion of Payday Lending Stores by Median Household Income and Census Tract



Creating Buffers around Bases using GIS

Table 1: Number of Payday Loan Stores around Bases

	5-mile Buffer	10-mile Buffer
Altus AFB	3	5
Camp Gruber		2
Ft Chaffee Maneuver Training Center		1
Ft Sill	23	27
McAlester Army Ammunition Plant	6	
Tinker AFB	43	151
Vance AFB	4	7
Total	73	199

Total payday stores in Oklahoma: 324

Kurban &Diagne, 2014.

Exhibit 3a

PUMA 3101 Real Cross-Tabulations

House Value ($)	Number of Children				
	0	1	2	3	4+
0–49,999	111	23	8	2	3
50,000–79,999	132	28	6	7	3
80,000–89,999	82	29	19	4	3
90,000–99,999	115	18	21	5	2
100,000–124,999	223	60	47	15	6
125,000–149,999	247	55	35	15	4
150,000–174,999	146	44	39	11	8
175,000–199,999	113	34	24	10	1
200,000–249,999	90	26	30	12	6
250,000–299,999	60	16	13	5	5
300,000–399,999	37	9	11	0	1
400,000–499,999	9	0	3	1	0
500,000+	16	2	3	1	0

PUMA = Public Use Microdata Area.

Exhibit 3c

IPF Cross-Tabulations for PUMA 3101*

House Value ($)	Number of Children				
	0	1	2	3	4+
0–49,999	106	24	10	4	3
50,000–79,999	124	28	12	8	4
80,000–89,999	94	21	17	4	2
90,000–99,999	115	22	17	5	2
100,000–124,999	234	58	42	11	5
125,000–149,999	238	55	41	14	8
150,000–174,999	158	40	33	11	5
175,000–199,999	113	32	26	9	3
200,000–249,999	93	29	29	9	5
250,000–299,999	58	18	14	6	3
300,000–399,999	30	11	10	5	2
400,000–499,999	7	3	2	1	1
500,000+	12	3	4	2	0

IPF = iterated proportional fitting. PUMA = Public Use Microdata Area.

** Values are rounded to the nearest whole number.*

Exhibit 3b

Hill-Climbing Cross-Tabulations for PUMA 3101*

House Value ($)	Number of Children				
	0	1	2	3	4+
0–49,999	110	24	7	2	4
50,000–79,999	128	29	8	8	3
80,000–89,999	94	20	21	2	0
90,000–99,999	116	21	15	8	1
100,000–124,999	236	63	35	12	5
125,000–149,999	243	48	43	14	8
150,000–174,999	153	39	40	11	5
175,000–199,999	109	35	30	5	3
200,000–249,999	98	24	25	11	6
250,000–299,999	55	21	13	6	4
300,000–399,999	24	13	12	6	3
400,000–499,999	3	4	4	2	0
500,000+	12	3	6	1	0

PUMA = Public Use Microdata Area.

** Values are rounded to the nearest whole number.*

References

- A Beginner's Guide to Creating Small Area Cross Tabulations, H Kurban, R Gallagher, GA Kurban, J Persky - Cityscape, 2011.
- Demographics of Payday Lending in Oklahoma, Haydar Kurban and Adji Diagne 2014.

http://coas.howard.edu/centeronraceandwealth/reports&publications/Oklahoma%20Payday%20Lending%20Report%20Final%20For%20Website.pdf