First Difference Analysis of the Motherhood Penalty in Vietnam: An Econometric Study Using VHLSS Data from 2004, 2006, and 2008

Pham Minh Tu

October 20, 2024

Abstract

This study examines the impact of having a small child on women's income in Vietnam using panel data from the Vietnam Household Living Standards Survey (VHLSS) between 2004, 2006, and 2008. By employing a First Difference model, we attempt to isolate the effect of motherhood status on income while controlling for age, education, and other factors. The findings indicate that mothers who take care of a small child experience a decrease in income by approximately 1.5%, as indicated by the coefficient $\beta = -0.015$. However, this effect is statistically insignificant, suggesting that the immediate impact of motherhood on income may not be as pronounced as expected within this sample and period.

Contents

1 Introduction

Despite increasing gender equality in many areas of life, women worldwide continue to face significant disadvantages in the labor market. One of the most well-documented challenges is the *motherhood penalty*, which refers to the negative impact that having children, particularly young ones, can have on a woman's career and earnings. This phenomenon has been extensively studied in developed countries, where researchers have demonstrated that mothers often face reduced wages, fewer promotions, and decreased employment opportunities compared to their childless peers or male counterparts. However, in many developing countries, the extent and nature of the motherhood penalty remain less explored.

Vietnam provides a unique context for studying the motherhood penalty due to its rapid economic growth and changing social landscape over the past few decades. Since the early 1990s, Vietnam has transitioned from a centrally planned economy to a market-oriented one, leading to substantial changes in labor market dynamics and household structures. The country has also experienced notable shifts in gender norms, with increasing participation of women in the labor force. However, traditional gender roles remain entrenched, particularly in rural areas, where women are often expected to balance both work and caregiving responsibilities. Understanding how these societal expectations influence women's economic outcomes is crucial for developing effective gender-sensitive policies in Vietnam.

The motherhood penalty is particularly relevant in Vietnam, where women are expected to perform the majority of household tasks, including child-rearing, even when they are employed full-time. This dual burden of work and family responsibilities can result in reduced productivity, missed work opportunities, or even withdrawal from the labor market. In this context, the decision to have a child can have long-lasting consequences for a woman's income trajectory and career prospects. Given that early childhood care is often a significant time demand, it is critical to understand how having a young child affects a woman's earnings.

The concept of the motherhood penalty typically encompasses several dimensions, including the direct impact on wages, career progression, and hours worked. It is important to distinguish between these different channels. For example, some women may reduce their working hours after becoming mothers, which leads to lower overall income, but their hourly wage rate may remain unaffected. Others may face discrimination in the workplace, where employers assume that mothers are less committed to their jobs, leading to lower wages and fewer promotions. In some cases, social expectations about gender roles can exacerbate the penalty, as women are often expected to prioritize family over career.

However, the empirical evidence on the magnitude of the motherhood penalty in developing countries is more limited compared to that of developed nations. One of the reasons for this is the scarcity of reliable longitudinal data, which is necessary to track the same individuals over time to observe how motherhood affects their earnings. This study leverages the rotating panel data from the Vietnam Household Living Standards Survey (VHLSS) from 2004, 2006, and 2008, which provides a unique opportunity to examine the effect of motherhood on income in Vietnam.

This study aims to fill a gap in the literature by investigating the motherhood penalty in Vietnam using data from the VHLSS. Specifically, it focuses on how the presence of a small child (under six years old) affects women's income. To isolate this effect, we employ a First Difference (FD) model, which helps control for unobserved individual-specific factors that might otherwise confound the analysis. By using this method, we can better estimate the causal impact of motherhood on women's earnings.

The main contribution of this research is twofold: First, it adds to the growing body of literature on gender inequality and labor market outcomes in developing countries by providing empirical evidence from Vietnam. Second, it offers insights into the specific mechanisms through which motherhood affects women's earnings, shedding light on the economic consequences of caregiving responsibilities in a rapidly changing society. Understanding these dynamics is essential for policymakers who are looking to promote gender equality and support women's economic empowerment.

1.1 Research Objective

The primary objective of this study is to quantify the effect of having a small child on women's income in Vietnam. Using the First Difference model, we estimate how changes in motherhood status (i.e., the presence of a young child) influence the income trajectories of Vietnamese women over the period 2004 to 2008. The research focuses on women who remained in the labor market after having children and examines how their income changed compared to when they were childless.

1.2 Significance of the Study

Studying the motherhood penalty in Vietnam is not only important for understanding gender inequality in the country but also for informing policies aimed at reducing disparities in the labor market. The findings from this research can help policymakers design interventions to support working mothers, such as more accessible childcare services, family-friendly work policies, or financial incentives for employers to provide flexible work arrangements. In the broader context, understanding the motherhood penalty can also inform debates on gender equality, poverty reduction, and economic development.

The findings of this study will contribute to the body of knowledge on how motherhood affects women's labor market outcomes in a developing country context, an area that is still underexplored in academic research. Furthermore, the insights gained from this research can serve as a foundation for future studies on the long-term effects of family structure on economic outcomes and labor market dynamics in Vietnam and similar economies.

2 Literature Review

The motherhood penalty refers to the income disparity faced by women who have children compared to those who do not. This phenomenon is widely studied across different contexts, with variations in findings due to factors such as geography, data, and methodologies. Two key papers provide valuable insights into the extent and mechanisms behind the motherhood penalty:

2.1 Is There a Motherhood Penalty? Decomposing the Family Wage Gap in Colombia (Gamboa & Zuluaga, 2013)

This paper explores whether mothers in Colombia experience a wage penalty and aims to decompose the family wage gap. The authors use a matching procedure developed by Ñopo (2008) as an alternative to the more common Blinder–Oaxaca decomposition method. This approach allows for a comparison between mothers and non-mothers with similar characteristics and addresses the differences in the distributions of individual characteristics between these two groups.

A major limitation of this study is the cross-sectional nature of the data, which prevents longitudinal analysis. Additionally, data from developing countries often lack the robustness seen in data from developed countries, which hinders certain types of analysis.

The study finds a wage gap of 1.73% between mothers and non-mothers in Colombia, significantly lower than previous estimates of 9.3% using traditional decomposition methods. Interestingly, when characteristics like schooling were accounted for, the unexplained portion of the gap diminished, showing no significant discrimination against mothers in the Colombian labor market [4]

2.2 The Motherhood Wage Penalty: A Meta-Analysis (Cukrowska-Torzewska & Matysiak, 2020)

This meta-analysis aggregates the findings from 39 studies to determine the average size of the motherhood wage penalty across different countries and tests various hypotheses regarding the persistence of the penalty. The authors perform a meta-analysis on 208 wage effects of having exactly one child and 245 wage effects related to the total number of children. They analyze different factors, including country-specific policies, social norms, and labor market conditions, that contribute to the penalty.

A key challenge in this study is the wide range of estimates across countries and periods, which makes it difficult to draw universal conclusions. In some countries, a significant portion of the gap could be attributed to public policies, while in others, discrimination or cultural norms played a larger role.

The meta-analysis finds that the average motherhood wage penalty is around 3.6–3.8%. The gap is smallest in Nordic countries, which have strong policies supporting work-family balance, and largest in Central and Eastern Europe and Anglo-Saxon countries, where work-family policies are weaker. The study finds that the penalty is mainly due to human capital depreciation during child-rearing periods, with some contribution from job sorting into lower-paying, family-friendly occupations [3].

2.3 Additional Literature

Other key studies have used methods such as fixed-effects models to control for unobserved heterogeneity and explore how motherhood impacts wages.

Studies from the U.S. and the U.K. frequently highlight how career interruptions and lower work effort contribute to wage penalties, and how certain policy interventions, such as parental leave and childcare provision, can mitigate the effects of motherhood on wages.

The existing literature highlights the complexity of the motherhood penalty, with the extent and causes varying significantly by country and policy environment. These findings underscore the need for country-specific research, as generalizations about the motherhood penalty are difficult to make.

Other key studies have used various methods, including fixed-effects models and experimental approaches, to explore how motherhood impacts wages in different contexts. For instance, [1] find that women in the U.S. face a significant wage penalty for motherhood, largely due to reduced work experience and employer discrimination. Similarly, [2] use experimental methods to show that mothers are perceived as less competent and less committed, which negatively affects their job prospects.

In a cross-national analysis, [6] emphasize the role of family policies and cultural attitudes in shaping the motherhood penalty, finding that countries with strong parental leave policies have smaller wage gaps. [7] focus on Denmark and show that while gender inequality is low before childbirth, it increases significantly afterward due to reduced labor market participation by mothers.

In Germany, [5] use trajectory modeling to demonstrate that the wage penalty for motherhood can persist for many years, further underscoring the long-term effects of childbearing on women's earnings.

3 Data Structure and Variable Construction

This study utilizes data from the Vietnam Household Living Standards Survey (VHLSS) for the years 2004, 2006, and 2008. The dataset was created by merging several rounds of surveys, focusing on the socio-economic characteristics of individuals and households. Specifically, data were sourced from the 'ho1', 'muc1', 'muc2', and 'muc4' datasets for each year, ensuring consistency in the variables across the different survey rounds.

3.1 Merging Procedure

To construct a balanced panel dataset, I first merged the data for each year individually by using common identifiers such as `tinh`, `huyen`, `xa`, `diaban`, `hoso`, and `matv`, which represent the province, district, commune, village, household, and individual ID, respectively. This step ensured that each individual could be tracked consistently across years. The merging process involved the following steps:

1. **Merging within each year**: For each survey year (2004, 2006, and 2008), I merged the 'ho1', 'muc1', 'muc2', and 'muc4' datasets by using the common identifiers.

2. **Creating the panel dataset**: After merging the data within each year, I proceeded to merge the datasets across years. However, due to inconsistencies and discrepancies in the 2002 data (where the `matv` variable was sometimes recorded differently for the same individuals in 2004), I decided not to include the 2002 data in the panel. Thus, the final panel dataset comprises individuals who participated in both 2004–2006 and 2006–2008, ensuring that each individual is present in at least two consecutive survey rounds.

During the merging process, I also checked for duplicate entries, removed rows with missing values, and converted certain variables from string to numeric formats to facilitate the analysis. Additionally, I renamed variables for clarity and readability.

3.2 Variable Selection and Construction

For the purposes of this study, I focused on a subset of variables that are relevant to understanding the relationship between motherhood and income. The key variables include socio-demographic characteristics such as age, education, marital status, and ethnicity, as well as income and employment-related variables. The columns retained in the dataset were as follows:

```
columns_to_keep = ['tinh', 'huyen', 'xa', 'diaban', 'hoso',
                   'ttnt', 'matv', 'm1ac1', 'm1ac1a', 'ttnt',
                   'schooling_years', 'dantoc', 'm1ac2',
                   'relation_to_head', 'm1ac4a', 'm1ac4b',
                   'm1ac5', 'm4ac1a', 'm4ac2', 'm4ac6',
                   'm4ac7', 'm4ac8', 'm4ac9', 'm4ac10b',
                   'total_income', 'm1ac6']
```

After retaining the desired columns, I further processed the data by computing and constructing new variables to enhance the analysis.

3.3 Construction of Key Variables

1. **Dependent Variable:** `log_income`
   The primary dependent variable in this study is the log of total income, which was calculated by taking the natural logarithm of the total income variable (`total_income`). The logarithmic transformation is commonly used to reduce the skewness of income data and interpret the coefficients in terms of percentage changes. For individuals with zero income, I dropped them from the analysis before applying the logarithmic transformation to avoid undefined values.

   $$\texttt{log\_income} = \log(\texttt{total\_income})$$

2. **Age and Age Squared**
   Age is another key variable in the analysis, as it captures the experience of the individual. In addition to using the actual age, I also constructed the square of the age (`age_sq`) to account for potential non-linear effects of age on income. This allows the model to capture diminishing returns to experience, which is common in wage regressions.

   $$\texttt{age\_sq} = (\texttt{age})^2$$

3. **Mother of a Small Child**
   The variable `mother_of_small_child` is the key independent variable of interest. It identifies whether a woman in the dataset is a mother with a small child under the age of three. This age threshold was chosen because it is expected that mothers of children younger than three years old might have less time to work due to caregiving responsibilities. The construction of this variable involved several steps:

   - First, I identified all females in the dataset using the variable `m1ac2` (where `m1ac2 == 2` indicates female).
   - Then, I identified all children under three years of age using the variable `m1ac5`.

- To identify mothers, I considered women who were either the head or wife of the household, or who were daughters in larger families where the child was the grandchild of the head.
- I merged the mother and child records and assigned the value of 1 to `mother_of_small_child` if the woman met the criteria, and 0 otherwise.

The code for this construction is provided below:

```python
females_panel      = panel_data[panel_data['m1ac2'] == 2]
children_panel     = panel_data[panel_data['m1ac5'] < 3]
mothers_head_wife = females_panel[females_panel['relation_to_head'].isin([1, 2])]
mothers_daughter   = females_panel[females_panel['relation_to_head'] == 3]
matched_head_wife = pd.merge(children_panel[children_panel['relation_to_head'] == 3], mothers_head_wife, on=['tinh', 'huyen', 'xa', 'diaban', 'hoso', 'year'], suffixes ('_child', '_mother'))
matched_daughters = pd.merge(children_panel[children_panel['relation_to_head'] == 6], mothers_daughter, on=['tinh', 'huyen', 'xa', 'diaban', 'hoso', 'year'], suffixes=('_child', '_mother')) all_mothers_panel = pd.concat([matched_head_wife, matched_daughters])

panel_data['mother_of_small_child'] = 0
panel_data.loc[panel_data[['tinh', 'huyen', 'xa', 'diaban', 'hoso', 'matv', 'year']].apply(tuple, axis=1).isin(all_mothers_panel[['tinh', 'huyen', 'xa', 'diaban', 'hoso', 'matv_mother', 'year']].apply(tuple, axis=1)), 'mother_of_small_child'] = 1


```

4. **Other Control Variables** The model also includes several control variables that help account for other factors affecting income:

   - `schooling_years`: This variable measures the number of years of education, ranging from 1 to 12.
   - `ethnic`: A dummy variable indicating whether the mother belongs to the ethnic majority (1 for Kinh, 0 for minority groups).

- marital_status: This is a binary variable indicating whether the individual is married (1 for married, 0 for otherwise).
- working?: A dummy variable indicating whether the individual is currently working in the labor market.
- canbo?: A binary variable showing whether the individual is a state employee or works in a government position.
- ttnt: This variable indicates whether the individual resides in an urban or rural area (1 for urban, 0 for rural).

3.4 Final Panel Data

After constructing the variables, I ensured that the dataset was consistent across the years 2004, 2006, and 2008. The final columns in the panel data are as follows:

```
columns_order = ['tinh', 'huyen', 'xa', 'diaban', 'hoso', '
    matv', 'm1ac1', 'm1ac1a', 'year', 'id','
    mother_of_small_child', 'log_income', 'age', 'age_sq', '
    schooling_years', 'ttnt', 'ethnic', 'marital_status', '
    working?', 'canbo?', 'total_income']
```

This structure allowed me to analyze how the dependent variable log income is affected by the presence of a small child in the household, while controlling for other important socio-economic factors. The inclusion of variables such as age, age_sq, and schooling_years allows the model to capture key aspects of human capital theory, where wage growth is a function of experience and education. Additionally, by including dummy variables like marital_status, working?, and canbo?, the model accounts for differences in labor market participation and employment type, which are crucial for understanding wage dynamics in Vietnam.

Moreover, the ethnic variable enables the analysis to consider the impact of being part of the ethnic minority, a factor that has historically influenced labor market outcomes in Vietnam. The ttnt variable, which distinguishes between urban and rural areas, helps control for geographic disparities in economic opportunities, as individuals in urban areas typically have better access to higher-paying jobs compared to their rural counterparts.

4 Model Selection and Application

In this section, I apply the *First-Difference Model* to estimate the impact of having a small child on the log of income (`log_income`). Our primary interest lies in the variable `mother_of_small_child`, and I control for several covariates such as age, schooling years, ethnicity, marital status, and whether the mother works in the public sector or not.

4.1 Handling of Time-Invariant Variables

The *First-Difference Model* is appropriate for this analysis as it eliminates time-invariant characteristics that may bias the estimates, such as location (urban vs. rural), ethnicity, and fixed personal traits. Unlike the fixed-effects model, which assumes constant time-invariant differences between individuals, the first-difference model effectively captures within-individual variations across two points in time.

The dataset contains observations from 2004 to 2008, but for the purposes of this analysis, I focus on individuals who are observed in two consecutive years (2004-2006, 2006-2008). Since the analysis only contains two-year spans and no specific treatment such as a policy change is identified, the *difference-in-difference* approach is unsuitable. The absence of a treatment or policy event further supports the choice of the first-difference model over difference-in-difference or fixed/random effects models.

Mathematically, the first-difference model can be expressed as:

$$\Delta Y_{it} = \beta \Delta X_{it} + \Delta \epsilon_{it}$$

Where ΔY_{it} is the change in the dependent variable (log of income) for individual i between two time periods, ΔX_{it} is the change in the independent variables for individual i, and $\Delta \epsilon_{it}$ is the change in the error term.

4.2 First-Difference Model Application

I first calculate the differences in the dependent and independent variables between the two time periods. The first difference model is then applied to the dataset:

```
X = panel_data[['mother_of_small_child', 'age', 'age_sq',
               'schooling_years', 'ttnt', 'ethnic',
               'marital_status', 'working?', 'canbo?']]
```

```
Y = panel_data['log_income']

first_diff = pd.concat([panel_data['id'], X], axis=1)
first_diff = first_diff.groupby('id').diff()

first_diff.dropna(inplace=True)
first_diff.reset_index(drop=True, inplace=True)
first_diff = first_diff.loc[:, (first_diff != 0).any(axis=0)]
del first_diff
```

The independent variables used include key demographic and work-related factors. I aim to assess how a mother's responsibility for a small child affects her income, controlling for these covariates. After taking the first difference, I remove the time-invariant variables and only keep variables that vary over time.

```
X_after = panel_data[['mother_of_small_child', 'age', 'age_sq',
                      'schooling_years', 'ethnic',
                      'marital_status', 'canbo?']]

panel_data.set_index(['id', 'year'], drop=False, inplace=True)

firstdf_regress = plm.FirstDifferenceOLS.from_formula(
    formula='log_income ~ mother_of_small_child + age + age_sq + ttnt + schooling_years + ethnic + marital_status + `canbo?`',
    data=panel_data
)
results_fd = firstdf_regress.fit(cov_type='clustered', cluster_entity=True)

table_fd = pd.DataFrame({
    'b': round(results_fd.params, 4),
    'se': round(results_fd.std_errors, 4),
    't': round(results_fd.tstats, 4),
    'pval': round(results_fd.pvalues, 4)
})
print(table_fd)
```

The `FirstDifferenceOLS` function is used to estimate the model, and the results (coefficients, standard errors, t-statistics, and p-values) are presented in the table. The clustering of standard errors is done at the individual level to account for within-individual correlation.

4.3 Justification for First-Difference Model

Given that our data spans only two years and does not contain a specific policy or event that could serve as a treatment, the *First-Difference Model* is more appropriate than a difference-in-difference model. The first-difference approach effectively controls for any unobserved time-invariant factors that may bias our results. Additionally, the fixed-effects or random-effects models could not capture the short-term within-individual changes as precisely as the first-difference model given the structure of our data.

5 Results and Findings

5.1 Data Overview

After the initial data processing, which involved merging the years 2004, 2006, and 2008, the dataset consisted of 189,580 observations and 31 variables. These variables covered a wide array of individual and household characteristics, such as demographic factors, education levels, employment status, and income details. Each individual was uniquely identified by the combination of `tinh`, `huyen`, `xa`, `diaban`, `hoso`, and `matv`, ensuring a consistent panel structure across the three years.

The primary focus of this study was on mothers with small children, defined as children under the age of three. This threshold was set based on the assumption that once a child reaches the age of three, the mother has more flexibility to work, as children can attend kindergarten, reducing the direct burden of childcare. To identify these mothers, we examined the household structure to determine which women were responsible for young children, including mothers who were either the head of the household, the wife of the head, or daughters living with parents whose children were the head's grandchildren.

During this process, all missing data (NaNs) and irrelevant or inconsistent observations were dropped. Moreover, I ensured that only the individuals observed in two consecutive years (2004-2006 and 2006-2008) remained in the dataset. Due to synchronization issues with the 2002 data, this year was excluded from the analysis. As a result of these procedures, the dataset shrank considerably, leaving 3,580 observations and 7 key variables for analysis. These variables included:

- `mother_of_small_child`: Whether the woman is responsible for a child under three years old,
- `log_income`: The natural logarithm of total income,
- `age`: The mother's age,
- `age_sq`: The square of the mother's age to account for non-linear effects,
- `schooling_years`: Number of years of schooling,
- `ethnic`: Whether the mother belongs to the ethnic majority or a minority,
- `canbo?`: Whether the mother is a public sector employee.

The reduction from 189,580 observations to 3,580 reflects the strict filtering process required to isolate the group of interest (mothers with small children) and ensure high-quality data for the regression analysis.

5.2 First-Difference Model Results

After constructing the final panel dataset, I applied the first-difference model to estimate the impact of having a small child on a mother's income, controlling for other variables such as age, schooling, and ethnicity. The results of this model are summarized in Table 1.

Table 1: First Difference Model Results

Variable	Coefficient	Std. Error	t-stat	p-value
mother_of_small_child	-0.015	0.0635	-0.2368	0.8128
age	0.098	0.0431	2.276	0.023
age_sq	0.0001	0.0005	0.172	0.8634
schooling_years	0.0037	0.018	0.2078	0.8354
ethnic	-0.0276	0.1022	-0.2705	0.7868
marital_status	0.1353	0.0901	1.5014	0.1334
canbo?	0.194	0.0402	4.8241	0.000

5.3 Discussion of Results

The first-difference model allows me to capture the within-individual changes over time, eliminating any time-invariant biases, such as ethnicity or rural/urban differences, which may confound the analysis. The main coefficient of interest is the variable `mother_of_small_child`, which represents the effect of having a child under three years old on a mother's income.

The estimated coefficient for `mother_of_small_child` is -0.015, which suggests that mothers responsible for small children experience a 1.5% reduction in income. However, the associated p-value of 0.8128 indicates that this effect is statistically insignificant. In other words, although the coefficient is negative, I do not have sufficient evidence to conclude that having a small child has a significant impact on a mother's income in the sample observed. This finding could suggest that while there may be an income reduction, other compensatory factors, such as support from family or the flexibility of employment, may mitigate the overall financial impact.

Several control variables show significant results. For example, the coefficient for `age` is 0.098, indicating that as the mother ages, her income increases. This is likely due to increased work experience and skill accumulation over time. The positive and significant coefficient on `canbo?` (0.194) suggests that mothers working in the public sector earn significantly higher incomes compared to those not employed in the public sector. This is consistent with the expectation that public sector employment offers more stable and often higher-paying jobs in Vietnam, especially for women.

Interestingly, variables such as `schooling_years` and `ethnic` are not statistically significant. This implies that, at least in this dataset, factors such as education and ethnicity may not have a strong short-term effect on the income changes of mothers with small children.

The coefficient for `ttnt` (urban/rural area) is negative, but not statistically significant, indicating that the urban/rural distinction might not play a crucial role in explaining income differences for mothers of small children in the short term. Similarly, `marital_status` shows a positive effect on income but is not statistically significant. This could imply that while being married may offer some financial stability, it does not translate into significant income changes for this group of mothers.

5.4 Interpretation of Results

The results from the first-difference model provide valuable insights into the dynamics of income for mothers with small children. While the main variable of interest, `mother_of_small_child`, does not appear to have a significant effect on income, other variables such as age and public sector employment do. This may reflect the broader economic environment in Vietnam during the 2004-2008 period, where job security and wage growth in the public sector were prominent features.

The insignificance of the motherhood penalty may suggest that, at least in the short term, Vietnamese mothers may benefit from family support networks or flexible work arrangements that allow them to maintain income levels despite the additional burden of childcare. Alternatively, it could be that the dataset does not capture the full extent of the income reduction experienced by mothers, as many may transition into part-time or informal work, which is not adequately measured.

Additionally, it is possible that the short time span between the two observed periods (two years) does not allow enough time for the full effects of motherhood on income to manifest. Future research could investigate longer time horizons or focus on specific sectors where the motherhood penalty may be more pronounced.

6 Conclusion and Policy Implications

6.1 Summary of Key Findings

This study finds that the motherhood penalty in Vietnam, as measured by the impact of having a small child on income, is relatively small and not statistically significant in the First Difference model. However, age and occupation status have significant effects on income.

7 Policy Implications

The findings of this study provide important insights for policymakers aiming to reduce the wage gap experienced by mothers with small children. The negative, albeit insignificant, impact of motherhood on wages calls attention to the structural challenges faced by women in the labor market, particularly when they are raising young children. Policymakers should consider implementing programs that support working mothers, such as more accessible and affordable childcare, extended parental leave, and flexible work arrangements. Such policies could help mitigate the motherhood penalty by allowing women to balance their work and family responsibilities more effectively.

In addition, targeted interventions, such as incentives for employers to provide childcare facilities or job protections for mothers returning to the workforce, may prove beneficial in improving the economic outcomes for mothers of young children. This study suggests that, even though the impact of motherhood on wages is not statistically significant in this case, the presence of a small negative coefficient highlights that some barriers remain, and further policy action may be required to ensure gender equity in wages.

8 Limitations and Future Research

While this study has provided valuable insights into the effects of motherhood on wages, there are several limitations that must be acknowledged. First, I encountered significant difficulties in handling the dataset because it was written in a special font used by the government. This required additional processing and effort to correctly import and read the data. Moreover, the data across different years were not optimally organized, leading to further complications in merging the datasets. Inconsistencies between years, particularly in the way individual identifiers and variables were recorded, posed a challenge, especially when trying to construct a consistent panel dataset.

Another limitation of this study is the exclusion of data from 2002 due to the high degree of inconsistency and differences in recording methods compared to later years. This decision reduced the number of time periods available for analysis, which may have affected the robustness of the findings. Future research should aim to address these issues by working with more standardized and harmonized data sources across different years. Additionally, incorporating longer periods of observation could provide more robust insights into the long-term effects of motherhood on income.

Furthermore, future studies could explore alternative models or treatments, such as incorporating policy interventions that explicitly target mothers with young children, to better understand the impact of these policies on labor market outcomes. Qualitative research may also complement quantitative analysis by investigating the lived experiences of mothers in the workforce, particularly in developing contexts like Vietnam.

9 Conclusion

This study has analyzed the impact of motherhood, particularly raising young children, on women's wages using the Vietnamese Household Living Standards Survey (VHLSS) data from 2004 to 2008. Through the application of the first-difference model, I aimed to estimate the effect of having a small child on the log of income, controlling for relevant factors such as age, education, ethnicity, and marital status.

Despite encountering several challenges in data processing and handling inconsistencies between years, the findings suggest that having a small child has a negative, but insignificant, effect on a mother's wages. These results underline the importance of further research in this area, as well as the need for better-organized datasets to allow for more accurate and comprehensive analysis. Moreover, the policy implications discussed highlight the potential benefits of supporting working mothers through policies that address the challenges they face in balancing work and family life.

The study contributes to the broader literature on the motherhood wage penalty and provides a basis for further exploration of gender wage gaps in Vietnam and other developing countries. Future research, with improved data quality and extended time periods, could offer deeper insights into the long-term effects of motherhood on income, thus better informing both policy and academic debates.

References

[1] Michelle J Budig and Paula England. The wage penalty for motherhood. *American Sociological Review*, 66(2):204–225, 2001.

[2] Shelley J Correll, Stephen Benard, and In Paik. Getting a job: Is there a motherhood penalty? *American Journal of Sociology*, 112(5):1297–1339, 2007.

[3] Ewa Cukrowska-Torzewska and Anna Matysiak. The motherhood wage penalty: A meta-analysis. *Social Science Research*, 88:102416, 2020.

[4] Luis Fernando Gamboa and Blanca Zuluaga. Is there a motherhood penalty? decomposing the family wage gap in colombia. *Revista Desarrollo y Sociedad*, 2013(72):169–197, 2013.

[5] Markus Gangl and Andrea Ziefle. Parenthood and earnings trajectories in germany: A group-based trajectory modeling approach. *Advances in Life Course Research*, 21:109–118, 2014.

[6] Joan R Kahn, Javier García-Manglano, and Suzanne M Bianchi. Childbearing, female employment, and the motherhood wage penalty in a cross-national perspective: The importance of work-family policies and cultural attitudes. *Social Science Research*, 45:1–17, 2014.

[7] Henrik Kleven, Camille Landais, and Jakob Egholt Søgaard. Children and gender inequality: Evidence from denmark. *American Economic Journal: Applied Economics*, 11(4):181–209, 2019.