Article

Estimating the Frisch Elasticity with New Instruments: Lessons from a Comparative Analysis

Muhammad Mahd Faisal[1]

[1] Muhammad Mahd Faisal (BSc Econometrics and Mathematics, 2022-2026), School of Economics, University of New South Wales, Australia; z5405472@ad.unsw.edu.au

How to cite
Faisal, M. (2025). Estimating the Frisch Elasticity with New Instruments: Lessons from a Comparative Analysis. *UCL Journal of Economics*, vol. 4 no. 1. DOI: 10.14324/111.444.2755-0877.2037

Peer review
This article has been peer-reviewed through the journal's standard double-blind peer review, where both the reviewers and authors are anonymised during review

Abstract
This paper estimates the Frisch elasticity of labour supply for men and women using US longitudinal microdata from the NLSY97 from 2007 to 2019. A central challenge in identifying labour supply elasticities is finding strong and valid instruments for wage changes. Building on existing approaches, initially, ASVAB cognitive test scores are used as an instrument. While ASVAB scores provide a valid instrument for men, they perform poorly for women. To address this, lagged income is proposed as a new instrument. Lagged income proves to be highly relevant and valid for both men and women, with first stage F-statistics exceeding conventional thresholds. Using this approach, the Frisch elasticity is estimated at approximately 0.45 for men and 0.73 for women, suggesting that women's labour supply is more responsive to wage changes. These results have important policy implications for the design of tax policy, welfare programmes, and labour market interventions. The findings also highlight the importance of carefully selecting instruments in empirical labour economics and open avenues for future research on sexspecific labour supply dynamics.

Keywords: Frisch Elasticity, instrument variables, cognitive ability, female labour supply, tax, 2SLS

1. Introduction

The Frisch elasticity measures the responsiveness of labour supply to predictable wage changes. The significant task here is to isolate the predictable change by removing the income effect and obtaining the pure substitution effect that is fundamental to understanding the labour market behaviour. Accurate estimation of this elasticity is crucial for informing public policy, including taxation, welfare programmes, stimulus programmes and macroeconomic modelling. It has far-reaching implications for the taxation policy as well (Keane, 2011).

However, estimation of the Frisch elasticity faces significant challenges. A major issue is the endogeneity of wages. The unobserved factors, such as uncertainty in individual productivity, preferences for leisure, or household decisions, may simultaneously influence both wages and hours worked, biasing ordinary least squares (OLS) estimates. As a result of this discrepancy, researchers have come up with a different approach. They have relied on running a two-stage least squares regression using an instrument variable for the change in log wages. Many in the past have used variables like parental education, an individual's years of schooling, and scores in cognitive tests as valid instruments. Yet, much of this literature focused solely on estimating Frisch elasticity of working-age men; therefore, the validity of these instruments was not always tested over diverse subgroups, such as by sex. Hence, a key focus in this paper is on identifying differences between sexes, as much of the economic literature highlights major distinctions between Frisch elasticities between men and women (Whalen & Reichling, 2012).

This paper revisits the estimation phase of Frisch elasticity by using a rich panel dataset, the National Longitudinal Survey of Youth 1997 (NLSY97). At first, a conventional instrument variable approach is used by setting ASVAB test scores as an instrument for the change in log wages. These ASVAB test scores are scores of a cognitive test conducted for most of the participants of the NLSY97 dataset in 1997. While ASVAB performs adequately for men, with evidence of strong relevance and exogeneity, it fails to serve as a strong instrument for women, leading to weak identification concerns. In response, we propose an alternative instrument, lagged annual income, as suggested by some prior economic literature and empirical precedents. We motivate and empirically assess the relevance and exclusion conditions for lagged income as it yields plausible Frisch elasticity estimates for both men and women.

Our findings highlight two key insights. First, the performance of standard instruments such as ASVAB is not uniform across both sexes, cautioning against blanket application without subgroup validation. Second, lagged income emerges as a promising, robust instrument capable of addressing some of the longstanding difficulties in labour supply estimation. The goal of this paper is to find credible estimates of Frisch elasticity by using various instruments and building upon the policy implications.

2. Literature Review

The estimation of the Frisch elasticity of labour supply has been a central concern in labour economics for decades. Hence, there is an extensive and rich economic literature already present for the estimation of Frisch elasticity. Early empirical work, most notably by MaCurdy (1980) and Altonji (1986), recognised the simultaneity problem between wages and hours worked and proposed the two-stage least squares (2SLS) estimation strategies. These studies often employed instruments such as parental background characteristics or cohort-level variables to isolate exogenous variation in wages. Despite these efforts, empirical estimates have varied considerably, with micro-level studies typically finding relatively small Frisch elasticities ranging from 0 to 0.5 for men, while the macroeconomic calibrations have ranged from 2 to 4 (Peterman, 2012).

Many papers over the years have used the NLSY97 data set for the purposes of their estimation and have subsequently used the ASVAB test scores as instruments. Test scores from the Armed Services Vocational Aptitude Battery (ASVAB) are believed to be credible instruments owing to their measurement being conducted during the subject's adolescence, making them largely predetermined and unaffected by short-run labour market shocks (Kniesner & Ziliak, 1999). While appealing in theory, the practical relevance of ASVAB as an instrument has sometimes been limited by weak first-stage relationships. This issue has existed for as long as Frisch elasticity has been estimated by economists.

Parental education has been proposed as another potential instrument, premised on its role in shaping educational attainment and, consequently, wages. However, the validity of this instrument has been debated as well. If parental background influences not just the productivity of an individual but also their labour supply decisions through networks and social norms, the exclusion restriction would be violated. Empirical studies such as Kahn (2007) and Goldin (2014) have emphasised the

https://doi.org/10.14324/111.444.2755-0877.2039

importance of carefully considering differences between both sexes in the determinants of labour market behaviour, which further complicates instrument validity across men and women.

A more recent body of work has highlighted the importance of distinguishing between predictable and unpredictable components of wage variation when estimating labour supply elasticities. Papers such as Blundell et al. (2008) and Low (2010) argue that predictable wage changes drive substitution effects without triggering strong income effects, making them ideal for identifying the Frisch elasticity. In line with this reasoning, lagged income emerges as a promising instrument: it captures predictable variation in economic circumstances while mitigating concerns about contemporaneous shocks or reverse causality.

Despite these advances, relatively few studies have systematically compared instrument validity for men and women or explored alternative instruments when conventional choices like education or cognitive ability fail. This paper contributes to the literature by critically evaluating the performance of traditional instruments, such as the ASVAB and parental education, separately for men and women, and by proposing lagged income as a theoretically grounded alternative. By doing so, it provides new evidence on the robustness of the instrument and sheds light on sex-specific dynamics in labour supply responsiveness.

3. Data and Description

This study uses microdata from the National Longitudinal Survey of Youth 1997 (NLSY97), a nationally representative panel of individuals born between 1980 and 1984 in the United States. The information on working hours and wages considered in this study is recorded from 2007 to 2019. The NLSY97 provides detailed information on labour market outcomes, demographics, educational attainment, and good proxy variables for judging the cognitive ability of participants. The analysis sample is constructed by focusing on annual labour income (including salaries, wages, and tips) and annual hours worked. See Appendix for detailed discussion on how certain variables were created.

Table 1: Summary Statistics for Key Variables (Men and Women)

	Men (Mean)	Women (Mean)	Difference Tests
Number of Observations	14,057	13,020	---
Annual Income	46,219	35,608	10,602 ***
Annual Hours Worked	2,213	1,934	279.00 ***
Wage	26.52	25.90	0.63 (ns)
Change in log (Wage)	0.127	0.132	-0.01 (ns)
Change in log (Hours Worked)	0.028	0.015	0.01 (ns)
ASVAB Score	5.12	5.41	-0.29 ***
Mother's Education (Years)	13.05	13.01	0.04 (ns)
Father's Education (Years)	12.95	12.99	-0.04 (ns)

Note: Difference = Mean(men) – Mean(women). Significance levels: * p<0.1, ** p<0.05, *** p<0.01, ns = not significant.

The table above provides summary statistics for the key variables used in this study, separately for men and women. In this paper, we use the term sex, following common convention in labour economics literature, when analysing differences between men and women. It is important to note that the NLSY97 only records sex at baseline (Round 1, 1997) with the two categories being "male" and "female." Therefore, all reported differences are based on the sex variable as recorded by the survey, and do not capture gender identity or variation beyond this binary classification.

On average, men worked approximately 2,213 hours per year, while women worked 1,934 hours, a difference of 279 hours that is statistically significant at the 1% level. Consistent with this, men also earned a significantly higher average annual income of $46,210 compared to $35,608 for women, with a highly significant income gap of $10,602. By contrast, the average hourly wages were statistically indistinguishable across sexes, 26.52 for men versus 25.90 for women, suggesting that the income disparity arises primarily from differences in total hours worked rather than wage rates. Turning to other variables, women scored slightly higher on the ASVAB test (5.41 versus 5.12), with this difference of –0.29 also significant at the 1%

https://doi.org/10.14324/111.444.2755-0877.2039

level. In our study, ASVAB has been rescaled by dividing it by 10,000 to ensure better interpretability of results. Parental education levels were very similar across sexes, with the differences being small and statistically insignificant. Changes in log wages and log hours also did not differ significantly between men and women.

These descriptive patterns provide an important backdrop for the regression analyses that follow. They highlight the importance of distinguishing hours from wages in explaining income differences and suggest that instrument performance may differ by sex. Given that everyone appears multiple times in the data, observations are not independent across years. Without adjustment, standard errors would be biased downward. Therefore, a cluster standard errors approach has been used at the individual level to allow for arbitrary correlation within each person over time. This technique accounts for serial correlation within individuals over time, ensuring that inference remains valid even when error terms are not independent across observations for the same person.

4. Methods and Model

It is a fundamental technique in the realm of econometrics to perform an OLS regression of the change of log hours on the change of log wages. This model (1) does include all the control variables, the independent variable and the dependent variables.

Model (1)

$$\Delta \ln(H_{it}) = \alpha_0 + \alpha_1 \, \Delta \ln(W_{it}) + \alpha_2 \, C_{it} + \epsilon_{it}$$

The OLS model has been displayed above. The Hit is the annual hours worked for the individual i at period t, Wit is the wage calculated using annual income and annual hours worked, and Cit is a vector of control variables. The control variables have been introduced in order to limit the influence of confounding and other irrelevant variables. These are year and race dummies that have been created to include year fixed effects and race fixed effects. Year fixed effects have been added to capture any changes in business cycles, and the race fixed effects have been added to account for any racial bias that may taint the estimation of the Frisch elasticity of labour supply. These controls partly mitigate concerns that cyclical economic fluctuations drive our estimates. Thus, while we do not directly include macroeconomic aggregates such as the unemployment rate in the baseline regressions, much of their effect is captured indirectly through the fixed effects structure. As discussed earlier, owing to the fundamental differences in labour market conditions for men and women, two separate models will be run.

However, the OLS regression fails to deliver credible estimates of Frisch elasticity because it cannot isolate the predictable changes in log wages from the unanticipated changes. To address this limitation, following the lead of classical labour economics papers like MaCurdy (1980) and Altonji (1986), a 2SLS regression is employed, using an instrument variable for the change in log wages. The instrument variable must satisfy the two basic conditions of relevance and exogeneity. Relevance requires the instrument to be correlated with the endogenous regressor, in this case, the change in log wages. This can be satisfied using the results of the first-stage regression. The instrument must be statistically significant according to the results of the firststage regression, and the F-statistic being higher than 10 gives it the credibility of being a strong regressor (Staiger & Stock, 1997). Exogeneity, on the other hand, requires that an instrument be uncorrelated with the error term of the main regression. Since there is only one instrument used, there are no statistical methods to test for the exogeneity of the instrument. Therefore, researchers typically must rely on a strong economic argument to pitch an instrument. The twostage least squares model is displayed below. ddjhdkf$_{sds}$

First stage: **Model (2)**

$$\Delta \ln(W_{it}) = \pi_0 + \pi_1 \, Z_{it} + \pi_2 \, C_{it} + \mu_{it}$$

Second stage: **Model (3)**

$$\Delta \ln(H_{it}) = \beta_0 + \beta_1 \, \Delta \hat{\ln}(W_{it}) + \beta_2 \, C_{it} + e_{it}$$

In the first stage, Zit represents the instrument variables, and in the second stage $\Delta \hat{\ln}(Wit)$ is the instrumented change. Choosing a valid instrument is critical to getting statistically significant and meaningful estimates of the Frisch elasticity of labour supply. Therefore, considerable research and review have been conducted to find a suitable instrument.

https://doi.org/10.14324/111.444.2755-0877.2039

4.1. Empirical Strategy

When selecting an instrument for the 2SLS model, the potential good candidates are those that have been used by other researchers in relevant micro studies. These variables include GPA of the individual, number of years of schooling, ACT scores, ASVAB scores, household income and parental education. In this paper, we have conducted two separate 2SLS regressions with two different instrument variables. One of them is ASVAB, which has been used in a few recent papers like Keane (2024), and the other is a newer yet old idea of using lagged income, which has never been used on its own as a single instrument variable; however, Altonji (1986) did pitch this as a very credible instrument. A more recent use of lagged income cautioned about serial correlation problems (Low, 2015).

For this paper, ASVAB was chosen as the main instrument. It satisfies the exogeneity condition almost seamlessly since it is administered prior to any wage changes and at a very young age for most participants in 1997. This ensures that it is not correlated with subsequent unpredictable shocks in wages and the absence of reverse causality. ASVAB test scores are assumed to be a proxy variable for an individual's ability. Ability impacts wages through its role in determining an individual's productivity, but we assume that once wages are accounted for, there is no direct effect of ability on hours worked. Hence, ability is only indirectly related to hours worked.

All the papers that use ASVAB as an instrument have only focused on estimating Frisch elasticity for working-aged men. Therefore, it was pertinent that we explore an instrument that has been used over the years to find Frisch elasticity of labour supply estimates for women. Many papers that have used parental education as an instrument for both men and women have found that it is a weaker instrument when compared to other available ones (Kniesner & Ziliak, 1999). A classical study into labour market dynamics of married women found that parental education can be a weak or even an invalid instrument for women owing to selection issues. The major issue that occurs is that relevance is often very weak, making the instrument either extremely weak or outright irrelevant. Therefore, a separate regression analysis with parental education as an instrument was not conducted. It is further discussed later.

This paper proposes using lagged income as an instrument to find the Frisch elasticity of men and women. This technique has not been used a lot in modern literature and was proposed by Altonji (1986), who used a variation of it with a combination of other instruments. Lagged annual income is plausibly exogenous since it is determined by factors from the previous period, such as past wages and economic conditions. These are predetermined relative to current-period labour supply shocks. Moreover, after controlling for year fixed effects and demographics through the race fixed effects, any correlation between lagged income and current unobserved determinants of labour supply is expected to be minimal. We recognise that labour markets may respond with lags of several quarters, however, since our data is annual/biennial, we argue that much of the very short-run inelasticity is already smoothed out at the yearly frequency. Hence, lagged income affects current labour supply decisions only indirectly through its correlation with the change in log wages. We further discuss the validity of this instrument in the limitations section later.

Much of the economic literature focusing on the estimation of Frisch elasticity of labour supply has addressed the problem of weak instruments. Therefore, in this paper, the Anderson-Rubin test has been used to examine the validity of weak instruments. This is because even a weak instrument can be a valid instrument.

https://doi.org/10.14324/111.444.2755-0877.2039

5. Results and Interpretations

Adopting the learnings from many previous research papers on this topic, it has become a standard practice at this point to run a simplistic OLS regression of change in log hours on change in log wages. The table below presents the OLS regression results of the change in log hours on change in log wages as displayed in model (1).

***Table 2**:* OLS Regression Results: Change in Log Hours Worked on Change in Log Wages

	OLS (Men)	OLS (Women)
R-squared	0.42	0.36
No. of Observations	14,057	13,020
Change in log wages Coefficient	-0.48*** (0.02)	-0.43*** (0.02)
F-statistic	80.19	74.82

Note: Robust standard errors are clustered at the individual level, provided in parentheses, rounded to two decimal places. Figures smaller than 0.005 appear as 0.00. Regression controls for year and race dummies. Significance levels: * p<0.1, ** p<0.05, *** p<0.01, ns = not significant.

The OLS coefficients being negative here is a big issue that cannot be ignored. The Frisch elasticity represents the responsiveness of labour supply to a change in wages. According to rational economic theory, the estimates of Frisch elasticity of labour supply cannot be negative. Therefore, we can conclude that using OLS for estimation is not a good strategy for either men or women.

The table below provides results for the first stage and second stage regression results for men and women estimated using two-stage least squares and ASVAB as the instrument. The first stage results are based on Model (2), while the second stage results are based on Model (3).

Table 3: 2SLS Estimates Using ASVAB Scores as Instrument (Men and Women)

	2SLS 1st Stage (Men)	2SLS 2nd Stage (Men)	2SLS 1st Stage (Women)	2SLS 2nd Stage (Women)
Dependent Variable	$\Delta\ln(W_{it})$	$\Delta\ln(H_{it})$	$\Delta\ln(W_{it})$	$\Delta\ln(H_{it})$
ASVAB Coefficient	0.01***(0.00)	-	0.00 (ns) (0.00)	--
F- statistic	11.54	--	7.83	--
$\ln(W_{it})$ Coefficient	--	0.58** (0.24)	--	1.00 (ns) (0.68)
Number of Observations	14,057	14,057	13,020	13,020

Note: Robust standard errors are clustered at the individual level, provided in parentheses, rounded to two decimal places. Figures smaller than 0.005 appear as 0.00. Regression controls for year and race dummies. Significance levels: * p<0.1, ** p<0.05, *** p<0.01, ns = not significant.

The results show a very different picture for both sexes. Starting with the men, the first-stage results show that ASVAB as an instrument is relevant. It has an F-statistic which is greater than 10, satisfying the Staiger-Stock condition (Staiger & Stock, 1997). In the second stage, it provides a positive value of Frisch elasticity, 0.58, which is also statistically significant at a 5% significance level. This is in line with most other micro study estimates that range from 0 to 0.55. These studies show that the range for women is typically higher than for men; however, in our case, the estimates for women are not reliable due to statistical insignificance. Delving further into this, the first-stage F-statistic for women is lower than 10, which gives the notion that ASVAB is a weak instrument. Moreover, the first-stage t-test for ASVAB is statistically insignificant at both 5% and 10% significance levels, showing that ASVAB has little to no predictive power in the case of a change in log wages. The

https://doi.org/10.14324/111.444.2755-0877.2039

second stage t-test for women also fails to reject the null hypothesis that the coefficient of change in log wages is statistically different from 0.

Since the ASVAB is a weaker instrument for women, the Anderson-Rubin test was conducted to see whether it is even a valid instrument for the estimation of Frisch elasticity of women. The table below shows the Anderson-Rubin test results for men and women to judge the validity of ASVAB as an instrument.

***Table 4**:* Anderson-Rubin Test Results for Validity of ASVAB Instrument (Men and Women)

	Men	Women
Dependent Variable	$\Delta\ln(H_{it})$	$\Delta\ln(H_{it})$
ASVAB Coefficient	0.00***(0.00)	0.00 (ns) (0.00)
F- statistic (ASVAB = 0)	14.48**	0.67 (ns)
Number of Observations	14,057	13,020

Note: Robust standard errors are clustered at the individual level, provided in parentheses, rounded to two decimal places. Figures smaller than 0.005 appear as 0.00. Regression controls for year and race dummies. Significance levels: * p<0.1, ** p<0.05, *** p<0.01, ns = not significant.

The Anderson-Rubin test was developed in 1949 and uses a very different approach to test whether our estimate of Frisch elasticity is significant or not. It is basically conducted through a reduced-form regression of the outcome of interest on the instrument itself, along with the control variables. The older approach has not been used a lot in the economic literature until it was shown that the AR test is more credible than the second-stage least squares t-test results (Keane, 2024). The AR test is recommended by theory if the instruments are weak, and it is as credible as the second stage least squares t-test when instruments are strong. However, in our case, nothing much changes as a result of the AR test. The ASVAB was a strong instrument for men, and according to the AR test, it remains a strong and valid instrument. Whereas, for women, both the 2SLS first stage and second stage, as well as the AR test, deduce that ASVAB is a weak and invalid instrument.

Again, we find ourselves stuck in the same cycle as many economists before us. The quandary is that the main instrument for this paper is valid for men and gives a realistic estimate of Frisch elasticity; however, it is weak and invalid for women. Although it was planned that parental education would be used as a supplementary instrument for change in log wages in the case of women, prior studies from Kniesner & Ziliak (1999) to Kahn (2007) warn us against this approach. Much of the traditional instruments have consistently underperformed in other studies as well, and that underscores the need for a better instrument.

Lagged annual income emerges as a natural and very powerful alternative. Income from the previous period is strongly correlated with the current wage rates due to wage persistence, satisfying the relevance condition in theory. At the same time, it is variable that is determined before the current period's labour supply decisions. This makes a strong case for the exogeneity of lagged income, since it is highly unlikely to be influenced by transitory shocks to labour supply that are specific to the current period. Moreover, lagged income is especially well-suited as an instrument for change in log wages for women, whose labour supply decisions may be influenced by short-term events like childbirth. Since these shocks tend to be localised in the current period, while last year's income reflects long-run earning capacity, the risk of reverse causality is lowered. Therefore, lagged income as an instrument offers a unique combination of theoretical logic and expected empirical strength. There is also evidence of the use of lagged income in some capacity in the works of the famed economist Joseph. G Altonji.

To maintain overall coherence between results and hold the analysis to the same standard for men and women, a two-stage least squares regression analysis is performed for both men and women using lagged income as an instrument for change in log wages. The table below shows the regression results for two-stage least squares based on Model (2) and Model (3) for both men and women using lagged income as a single instrument.

https://doi.org/10.14324/111.444.2755-0877.2039

***Table 5**:* 2SLS Estimates Using Lagged Income as Instrument (Men and Women)

	2SLS 1st Stage (Men)	2SLS 2nd Stage (Men)	2SLS 1st Stage (Women)	2SLS 2nd Stage (Women)
Dependent Variable	$\Delta\ln(W_{it})$	$\Delta\ln(H_{it})$	$\Delta\ln(W_{it})$	$\Delta\ln(H_{it})$
Lagged Income Coefficient	-0.00***(0.00)	-	-0.00***(0.00)	--
F- statistic	21.13	--	12.00	--
$\ln(W_{it})$ Coefficient	--	0.45 (0.06)	--	(0.73) (0.10)
Number of Observations	11,694	11,694	10,735	10,735

Note: Robust standard errors are clustered at the individual level, provided in parentheses, rounded to two decimal places. Figures smaller than 0.005 appear as 0.00. Regression controls for year and race dummies. Significance levels: * p<0.1, ** p<0.05, *** p<0.01, ns = not significant.

Starting with the interpretation of results for men, they are comparable to those achieved using the ASVAB as the instrument. The first-stage regression yields a F-statistic greater than 10, providing evidence of a strong instrument, and the first-stage t-test indicates that lagged income is a statistically significant predictor of change in log wages, even at a 1% significance level. The Frisch elasticity estimate for men is 0.45, in line with micro studies, and it is statistically significant based on the second-stage t-test. The value shows that a 1% increase in wage would lead to a 0.45% increase in hours worked by men.

Now moving on to the analysis for women. This is where the results are surprising relative to ones achieved using ASVAB as an instrument. The first-stage t-tests of lagged income show that it is statistically significant even on a 1% significance level, and the F-statistic is also greater than 10, both in contrast with what was observed with ASVAB as an instrument.

Moreover, the second stage regression yields a statistically significant value of Frisch elasticity, 0.73, based on the second stage t-tests. Such a figure entails that a 1% increase in wage leads to a 0.73% increase in hours worked, all else held constant. This estimate, although a little higher than what modern studies support, is in line with their expectation of the Frisch elasticity for women being higher than men. The latter part of the argument is also something that is debated often. Classical economic literature, like Altonji (1986) and Kniesner & Ziliak (1999), suggests that the Frisch elasticity for men is usually higher than for women. However, modern studies estimate the Frisch elasticity of women to be higher than men (Blundell et al., 2016b). This shift between the studies from two different times can also be attributed to rapidly altering labour market dynamics, since women in the past had their labour supply decision constrained by familial roles and societal norms. Today, women's labour supply decisions are increasingly shaped by wage incentives, as institutional and social constraints on hours worked have eased.

These Frisch elasticity values of 0.45 for men and 0.73 for women not only show individual behaviour but also reflect a deeper structural transformation, like women's work and economic participation. The NLSY97 dataset contains data for working women from 2007 to 2019, a time of broader shifts in the labour market. During this period, women's labour force participation strengthened significantly, with more women pursuing a continuous career trajectory, attaining higher education, and becoming primary or joint household earners (Hurst, 2014). As a result, modern working women are more sensitive to wage changes, while men typically are more rigid in this regard. Interestingly, the shift in women's Frisch elasticity from the times of classical labour economics micro studies to modern labour economics coincides with increased workplace flexibility, higher access to childcare and evolving social norms

5.1. Policy and Research Implications

Now we have the values of Frisch elasticity for men and women. We have used sound reasoning to explain the reasons behind Frisch elasticity for women being higher than men, also supported by modern economic literature. However, these numbers are not just digits that show a clearer picture of the labour supply decisions of individuals. Instead, they have wide-ranging policy implications. A higher Frisch elasticity implies that workers are more likely to adjust their labour supply when incentivised financially. Consequently, wage-based policies such as reductions in marginal income tax rates, wage subsidies, or earned tax credits are likely to have major effects on aggregate labour supply and productivity, especially on women.

Previous research has established a strong link between the Frisch elasticity of labour supply and the costs of taxation. A comprehensive study into the differences in labour supply across countries, specifically Europe and the United States, could largely be explained by differences in tax rates (Prescott, 2004). The paper argued that higher rates of taxes lead to lower labour supply through substitution effects. Similarly, Rogerson & Wallenius (2009) presented that the estimates of Frisch elasticity are critical for evaluating the welfare costs of taxation. The paper showed that higher elasticities imply larger distortions and greater deadweight loss from taxation. Since this experiment was from the data collected in the United States, it could be replicated in future for developing countries. Major emerging economies are witnessing record female labour participation rates, and their Frisch elasticity being higher is a huge advantage for the growing economies since they tend to have higher levels of wage growth each year.

The difference in Frisch elasticity observed here between men and women also points to the importance of modelling labour supply behaviour in a heterogeneous framework. Static models that assume uniform elasticity across individuals may misrepresent aggregate responses to policy interventions. Dynamic labour models, such as the ones proposed by Imai (2004), suggest that labour supply decisions of individuals are greatly influenced by intertemporal considerations. Individuals are effectively trading off labour supply today against future leisure and consumption. Such life-cycle models can be built for women as well to observe sexspecific dynamics, greatly improving the predictive power of the estimates. The change of Frisch elasticity of women from being estimated as lower than that of men to being estimated as higher than that of men is also an interesting phenomenon. A lot of this can be attributed to the change in society with regard to working hours for women, the removal of institutional barriers, and government support through childcare benefits. Guner et al. (2020) emphasise that child-related transfers have sizable effects on women's labour supply, reflecting their disproportionate childcare responsibilities. This institutional channel contributes to higher observed elasticities for women in recent studies. While our analysis does not explicitly model family policy, it reinforces the interpretation that sex differences in Frisch elasticities are shaped by social policy and household structure as much as by intrinsic responsiveness to wages. As shown by Blundell et al. (2016a), the labour supply of women is especially responsive to shocks, consistent with the higher elasticity results. These insights motivate our sex-disaggregated estimates and inform our discussion of the validity.

Carrying on from this, there are several avenues of research. The use of lagged income as a strong instrument across sexes could be further validated across multiple datasets. Moreover, future work in this realm could explore life-cycle patterns of Frisch elasticity, investigating whether the responsiveness to wages changes over the working life of an individual.

5.2. Limitations

This is a comprehensive study into a very broad subject which has captivated the attention of economists for many decades, and it is also a topic of debate between micro-economists and macro-economists. Some of the assumptions we made in the course of this study have also left our estimate vulnerable to open debate. Therefore, the limitations of this study need to be discussed in detail.

One limitation of our specification is the omission of local labour market conditions, such as state-level unemployment rates. While year fixed effects capture national-level business cycle fluctuations, they do not capture cross-sectional heterogeneity in regional labour demand. The omission of state-level unemployment risks residual confounding from regional cycles. Future work could incorporate such measures to better isolate the role of macroeconomic shocks in shaping labour supply responses. It could also feature explicit measures of time preferences. Individuals with higher patience or lower discount rates may respond differently to wage changes, and this heterogeneity could influence our elasticity estimates. While our instruments aim to capture exogenous variation in wages, unobserved heterogeneity in time preferences remains a possible source of bias. The use of lagged income as an instrument is also a point of contention. As discussed above as well, many papers like Low (2015) have warned about issues regarding serial correlation when using lagged income as an instrument. Although in our case the use of annual/biennial data mitigates within-year adjustment concerns, if wage or preference shocks persist over multiple years, lagged income could remain correlated with contemporaneous hours, weakening the exclusion restriction.

Finally, future research could integrate measures of non-wage job amenities like remote work options into labour supply models, especially for women. In a technologically developing atmosphere with greater use of remote work options, non-wage benefits increasingly constitute a bigger part of the compensation than they have traditionally. The advent of AI and its rapidly increasing use have transformed occupational tasks, job security, and skill requirements. It is predicted that GenAI will disproportionately impact professional and clerical roles where women are often overrepresented. If GenAI raises

schedule flexibility in women-dominated professional roles, Frisch elasticities may rise via lower adjustment costs. On the contrary, if it compresses hours through automation, elasticities could fall, limiting the external validity of 2007-2019 estimates to future periods. Similar patterns could be expected for Frisch elasticity of men depending on how wages increase/decrease. These transformations mean that the gender-specific patterns we identify may not generalise fully to the future labour markets. Our findings should therefore be interpreted as evidence specific to the context under study, while recognising that the underlying mechanisms of labour supply adjustment may evolve over time. To sum it up, this change would not be something completely new. As we discussed above in the paper, the labour supply elasticities of men and women changed based on historical contexts; hence, their alterations in the future are normal and expected.

5.3. Conclusion

The paper estimates Frisch elasticity of labour supply for men and women using longitudinal data from NLSY97 and addresses longstanding challenges like weak instrument issues, as well as empirical identification of labour supply responsiveness. Building upon traditional instruments such as ASVAB and parental education, the paper proposes the use of lagged income as an instrument. While ASVAB does reasonably well for men, it performs poorly for women. Using lagged income as a new instrument, reliable and logically sensible values of Frisch elasticity for men and women were generated, 0.45 and 0.73, respectively.

The findings highlight the critical importance of instrument choice in empirical labour economics and suggest that sex-specific heterogeneity must be incorporated into models of labour supply behaviour. Future research could further investigate various topics as discussed above. There needs to be more modern papers investigating the validity of lagged income as an instrument. Until further research, this topic remains open and debatable.

References

Altonji, J.G., 1986. Intertemporal substitution in labor supply: Evidence from micro data. Journal of Political Economy, 94(3, Part 2), pp.S176–S215.

Blundell, R., Dias, M.C., Meghir, C. and Shaw, J., 2016a. Female labor supply, human capital, and welfare reform. Econometrica, 84(5), pp.1705–1753.

Blundell, R., Pistaferri, L. and Preston, I., 2008. Consumption inequality and partial insurance. American Economic Review, 98(5), pp.1887–1921.

Blundell, R., Pistaferri, L. and Saporta-Eksten, I., 2016b. Consumption inequality and family labor supply. American Economic Review, 106(2), pp.387–435.

Goldin, C., 2014. A grand gender convergence: Its last chapter. American Economic Review, 104(4), pp.1091–1119.

Guner, N., Kaygusuz, R. and Ventura, G., 2020. Child-related transfers, household labour supply, and welfare. Review of Economic Studies, 87(6), pp.2290–2321.

Hurst, K., 2024. U.S. women are outpacing men in college completion, including in every major racial and ethnic group. Pew Research Center. Available at:
https://www.pewresearch.org/short-reads/2024/11/18/us-women-are-outpacing-men-incollege-completion-including-in-every-major-racial-and-ethnic-group [Accessed 22 August 2025]

Imai, S. and Keane, M.P., 2004. Intertemporal labor supply and human capital accumulation. International Economic Review, 45(2), pp.601–641.

Kahn, L.M., 2007. The gender pay gap: Have women gone as far as they can? Academy of Management Perspectives, 21(1), pp.7–23.

Keane, M.P., 2011. Labor supply and taxes: A survey. Journal of Economic Literature, 49(4), pp.961–1075.

Keane, N., 2024. Robust inference for the Frisch labor supply elasticity. Johns Hopkins Carey Business School Research Paper (forthcoming), UNSW Business School Research Paper (forthcoming).

Kniesner, T.J. and Ziliak, J.P., 1999. Estimating life cycle labor supply tax effects. Journal of Political Economy, 107(2), pp.326–359.

Low, H. and Pistaferri, L., 2010. Wage risk and employment risk over the life cycle. American Economic Review, 100(4), pp.1432–1467.

Low, H., 2015. Disability insurance and the dynamics of the incentive–insurance trade-off. American Economic Review, 105(10), pp.2986–3029.

MaCurdy, T.E., 1980. An empirical model of labor supply in a life cycle setting. NBER Working Paper No. 421.

Peterman, W.B., 2012. Reconciling micro and macro estimates of the Frisch labor supply elasticity. Finance and Economics Discussion Series, Federal Reserve Board of Governors.

Prescott, E.C., 2004. Why do Americans work so much more than Europeans? NBER Working Paper No. 10316.

Rogerson, R. and Wallenius, J., 2009. Micro and macro elasticities in a life cycle model with taxes. Journal of Economic Theory, 144(6), pp.2277–2292.

Staiger, D. and Stock, J.H., 1997. Instrumental variables regression with weak instruments. Econometrica, 65(3), pp.557–586.

Whalen, C. and Reichling, F., 2012. Review of estimates of the Frisch elasticity of labor supply. Congressional Budget Office Working Paper 2012-13.

https://doi.org/10.14324/111.444.2755-0877.2039

Appendix

1. Hourly wages are computed by creating a ratio of annual income by annual hours worked. The natural logarithm of wages and hours is taken, and then first differences are calculated to obtain changes in log wages and changes in log hours worked between survey years. These changes form the basis of the entire process to calculate Frisch elasticity.

2. ASVAB variable was rescaled to ASVAB divided by 10,000 for the estimates to be more interpretable. Robust standard errors are clustered at the individual level, provided in parentheses in each table below the coefficient. All figures are rounded to two decimal places. Figures smaller than 0.005 appear as 0.00. Regression controls for year and race dummies are included in all regressions. Significance levels are displayed using commonly accepted practice: * p<0.1, ** p<0.05, *** p<0.01, ns = not significant.

Figure 1: Distribution of Lagged Income (Men). Histogram showing the distribution of lagged income for men, NLSY97 sample.

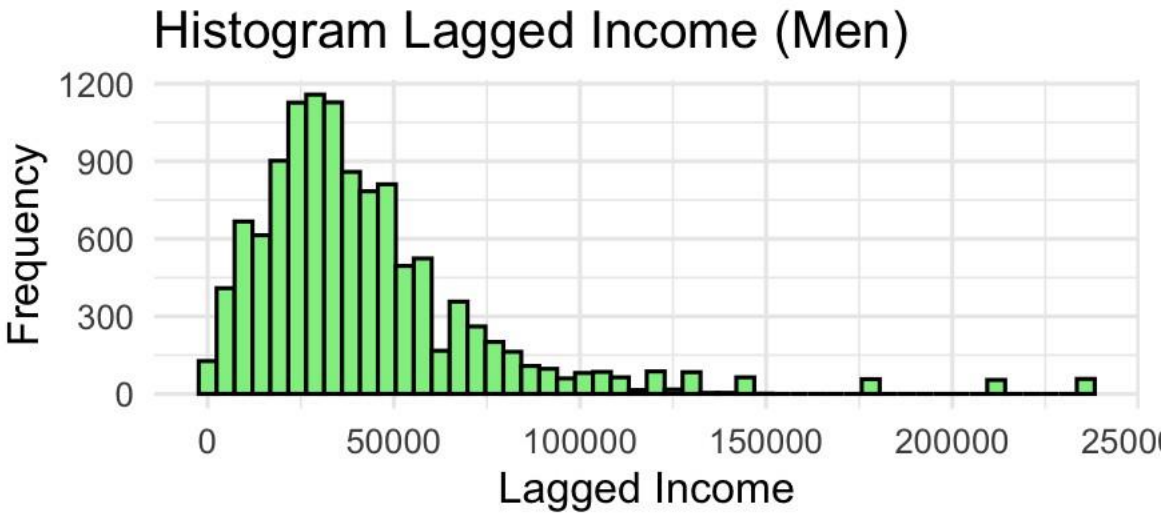
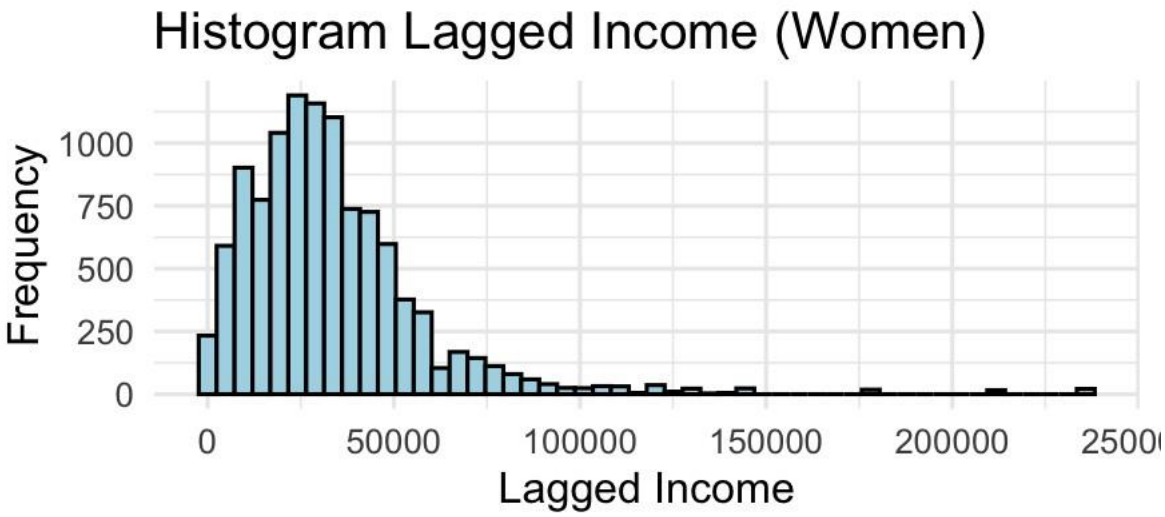


Figure 2: Distribution of Lagged Income (Women). Histogram showing the distribution of lagged income for women, NLSY97 sample.



3. To examine the distribution of the lagged income variable used as an instrument, histograms were generated separately for men and women. The histogram for men shows a right-skewed distribution, with most observations concentrated between US$20,000 and US$60,000 of lagged income. For women, the distribution is similarly right-skewed but slightly more

concentrated at lower income levels compared to men. This pattern is consistent with the summary statistics reported earlier in the paper.

4. Regarding Data and Code Availability, the analysis in this paper was conducted using Stata 16 for regression analysis and R (version 4.2.2) for cleaning data and generating plots. Replication code is available from the author upon reasonable request. The dataset used (NLSY97) is publicly available from the U.S. Bureau of Labor Statistics at the NLS Investigator.

5. We complement conventional 2SLS inference with weak-instrument-robust tests. AR inference remains valid even when the first stage is weak, so significance is not driven by large first-stage F-statistics alone.

Table 6: Anderson-Rubin Test Results for Validity of Lagged Income Instrument (Men and Women)

	Men	Women
Dependent Variable	$\Delta\ln(H_{it})$	$\Delta\ln(H_{it})$
ASVAB Coefficient	0.00***(0.00)	0.00***(0.00)
F- statistic (ASVAB = 0)	108.70**	133.76**
Number of Observations	11,694	10,735

Note: Robust standard errors are clustered at the individual level, provided in parentheses, rounded to two decimal places. Figures smaller than 0.005 appear as 0.00. Regression controls for year and race dummies. Significance levels: * p<0.1, ** p<0.05, *** p<0.01, ns = not significant.

6. The above table (6) shows the results of an AR test for lagged income as the sole instrument. The AR results strongly suggest the validity of the instrument for both men and women.

https://doi.org/10.14324/111.444.2755-0877.2039