Online Appendix
Trade, Maternal Time Costs, and Sex Selection: Evidence from Vietnam

Nghiêm Q. Huỳnh   Ngoc Nguyen

A Family Interrelationship Algorithm

Variables categorizing family interrelationships within the household are crucial in this paper. In particular, we need "pointer" variables identifying each person's mother, father, and spouse. VHLSS has a variable that classifies each individual's relationship to the household head. However, this variable is ambiguous and erroneous for the following reasons. First, extended family is common in Vietnam: roughly 34.91% of the VHLSS sample lives in the same household with relatives other than their nuclear family members. Because VHLSS does not identify subfamilies within the extended household, the "relate-to-head" variable will miss the interrelationship of subfamilies other than the head's. Second, with the exception of VHLSS 2002, all other waves do not identify "children-in-law." An immediate consequence is missing spousal links and therefore parental links in many cases. Thus, we write an algorithm to improve the precision of parent-child and spousal pairings in VHLSS.

We adapt a similar algorithm that has been used by the Minnesota Population Center (Sobek & Kennedy, 2009) to generate "pointer" variables for IPUMS–International Census Data. We utilize the four variables: relationship to household head, age, and marital status, in combination with the relative position of household members in the roster listing, to infer relationships. We first establish spousal links (generate variable SPLOC); then we find the mother and father for each individual (generate variables MOMLOC and POPLOC). Regarding parental linkage, we look for the mother before looking for the father. Once we find the mother for a child, we assign her husband as the father of this child. Only when we cannot find a child's mother do we locate its father (and assign his wife as the mother).

Tables A1 and A2 document the rules we apply to these matching tasks. Our algorithm applies the rules sequentially: if the first rule finds a match (spousal match or parental matches) for a given individual, the second rule no longer applies to this person, and so on. Whenever there are ambiguous multiple potential spouses or multiple potential parents, we drop the entire household from the sample.

Fortunately, VHLSS 2014 and 2016 provide two variables to locate the biological fathers and mothers of children under 16 years old. We use this information to test our algorithm and report the results in Table A3. We find the correct mother for 95.27% of these children and the correct father for 90.99% of them. The algorithm also finds both parents correctly for 88.95% of the children. The correction rate for both biological parents is lower than that for each biological parent because our algorithm also counts stepmothers and stepfathers. With these results, we are very confident that our algorithm does a good job at identifying family interrelationships.

Table A1: Rules for SPLOC Construction

Rule	**Individual's relationship to head**	**Partner's relationship to head**	**Age difference**	**Both Married**	**Require adjacency**	**Only applicable to 2002**	**Notes**
Strong couple pairing, couple adjacency preferred							
	Head	Spouse	No	No	No		1
	Parent	Parent	No	Yes	Yes		1
	Grandparent	Grandparent	No	Yes	Yes		1
	Child	Child-in-law	No	Yes	Yes	Yes	1
Weak couple pairing, couple adjacent							
	Grandchild	Grandchild	Yes	Yes	Yes		1, 2
	Other relationship	Other relationship	Yes	Yes	Yes		1, 2
	Sibling	Sibling	Yes	Yes	Yes	Yes	1, 2
	Grandchild	Other relationship	Yes	Yes	Yes		1, 2
	Sibling	Other relationship	Yes	Yes	Yes	Yes	1, 2
	Child	Other relationship	Yes	Yes	Yes		1, 2
Weak couple pairing, special type child-child							
	Child	Child	Yes	Yes	Yes		1, 2, 6
Weak couple pairing, couple not adjacent							
	Child	Child-in-law	No	Yes	Closest proximity	Yes	5
	Child	Other relationship	Yes	Yes	No		2, 3
	Child	Child	Yes	Yes	No		1, 2, 4, 6

Notes:

1. Drop the entire household if there is any person that could be assigned to 2 couples by the adjacency rule.
2. A woman can be no more than 20 years older or 35 years younger than a potential male partner.
3. For non-adjacent couple pairing, among the potential spouses who satisfy the age and marital requirements, select the person who is closest in age and impose that the husband is older than the wife.
4. For child-child non-adjacent couple pairing, drop all households where this rule yields multiple potential spouses.
5. These non-adjacent couples are matched based on having closest proximity to each other.
6. Once the couple is identified, the in-law is distinguished. For child-child couples, assume the first listed spouse is the biological child to the household head, hence the second listed spouse is the child-in-law.

B Model Details

We extend the quantity-quality model of Jones et al. (2011) to incorporate son preference. The model has two distinguishing features that reflect the Vietnamese context. First, households desire to have at least one son to continue the family lineage. Second, mothers bear the sole responsibility for child-rearing while also participating in the labor market.

Table A2: Rules for Child-Parent Relationship Construction (after SPLOC is Generated)

Rule	**Child's relationship to head**	**Parent's relationship to head**	**Age difference**	**Proximity requirement**	**Only applicable to 2002**	**Notes**
Links involving Head, Spouse, and Grandparent (unambiguous)						
	Child	Head, spouse	No	No		
	Child-in-law	Head, spouse	No	No	Yes	
	Head	Parent	No	No		
	Spouse	Parent	No	No		
	Sibling	Parent	No	No	Yes	
	Parent	Grandparent	No	No		
Links between grandchildren and children						
	Grandchild	Child, child-in-law	15–44	Weak		1
	Grandchild	Other relationship	15–44	Weak		1, 2
Links involving other relatives						
	Other relationship	Grandchild	15–44	Weak		1
	Other relationship	Other relative	15–44	Weak		1
	Other relationship	Sibling	15–44	Weak		1
	Other relationship	Child, child-in-law	15–44	Weak		1, 3

Notes:

1. Weak proximity requires that the child must be listed after its potential mothers (or potential fathers if it has no potential mothers); among them, its mother is the one listed closest to the child.

2. Impose that no person with code "Child" is present in the household. The mother of the grandchildren in these cases tend to be listed as "Others" since there is no category for "Child-in-law."

3. Impose that no person with code "Grandchild" is present in the household. These cases tend to mix up the numerical code for "Grandchild" (code 6) and that for "Other relationship" (code 7).

We incorporate these assumptions into a standard Q-Q model, wherein mothers assume the primary caregiver role. Specifically, let's consider a household comprising a woman of reproductive age and her husband. Referring to them as a mom (m) and dad (d), they jointly determine their private consumption c_g, leisure ℓ_g, where $g \in \{m, d\}$, the number of children n, and the quality of their children q, with $q = 1$ indicating a son and 0 otherwise.[1]

Alongside interpreting quality as the child's sex, we also establish the household's desire to have at least one son, which we denote as $Q \equiv qn$, representing the effective number of sons. In the context of logarithmic utility assumption, the model implies that a household would prefer having a son ($Q = 1$) over, for instance, having three daughters ($n = 3$) with no sons ($q = 0$). The utility function specific to each parent g is defined

[1] We actually let q be a continuous variable in the unit interval $[0, 1]$. All variables n and Q here are continuous to allow simple comparative statics.

Table A3: Performance of our algorithm in locating parents

	Freq.	Percent	Cum.
Both mom and dad correct	80021	88.95	88.95
Dad correct, mom incorrect	1831	2.04	90.99
Dad incorrect, mom correct	3853	4.28	95.27
Both mom and dad incorrect	4253	4.73	100.00
Total	89958	100.00	

Notes: This table summarizes the comparison between the parent locators generated by our algorithm with the true parents locators provided by VHLSS 2014-2016 for children under 16 years old.

as follows:

$$U_g = \alpha_c \log(c_g) + \alpha_\ell \log(\ell_g) + \alpha_n \log(n) + \alpha_q \log(Q),$$

where $g \in \{m, d\}$, and α_i represents the weight of preference assigned to the corresponding component, including consumption, leisure, fertility, and the effective number of sons.

While both parents have the same amount of time available to allocate between work and leisure, it is only the women who take care of the children, incurring a time cost of γ per child. Additionally, the household incurs a cost of p_q for the effective quality of the children Q. This cost encompasses the overall expenses associated with sex-selective abortion, including economic, physical, and psychological factors, as well as the benefits it brings to the couple. An example of p_q could be the relative returns of daughters compared to sons, which parents may forecast based on the current relative returns of women's work in the labor market.

In addition to labor incomes, the household also receives non-labor incomes denoted as I. This includes transfers from other members of the household or the value of their land, as discussed in Almond et al. (2019). Furthermore, children and their quality are considered public goods within the household. The optimization problem for the household can be formulated as follows:

$$\max_{\{c_m, c_d, n, q\}} \lambda_d U_d + \lambda_m U_m$$
$$\text{s.t. } c_m + c_d + p_q Q \leq I + w_d(1 - \ell_d) + w_m (1 - \ell_m - \gamma n).$$

where λ_m represents the bargaining weight of the mother, λ_d represents that of the father, and $\lambda_m + \lambda_d = 1$.

By defining the household's total income as $W = I + w_m + w_d$, we can derive the following solutions after taking logs:

$$\log q = \log(\gamma \alpha_q / \alpha_n) - \log p_q + \log w_m,$$
$$\log \ell_m = \log(\lambda_m \alpha_\ell) + \log W - \log w_m,$$
$$\log n = \log(\alpha_n / \gamma) + \log W - \log w_m.$$

These equations generate twelve comparative static predictions summarized in Table 1 of the main text.

The model's central predictions stem from three features. First, the childcare burden is asymmetric, where only mothers spend time on childcare, making their wages uniquely affect the quantity-quality trade-off. Second, parents value having sons ($\alpha_q > 0$), creating demand for sex selection. Third, changes in paternal wages and non-labor income affect fertility through pure income effects, without creating work-childcare tensions. These mechanisms generate the distinct predictions that differentiate our model from alternatives, as summarized in the main text.

B.1 Bargaining Extension

We now extend our model to incorporate household bargaining. Each parent's utility function includes gender-specific preferences for both fertility and sons:

$$U_g = \alpha_c \log(c_g) + \alpha_\ell \log(\ell_g) + \alpha_{ng} \log(n) + \alpha_{qg} \log(Q)$$

The household maximizes a weighted sum of utilities where bargaining weights depend on relative wages:

$$\max_{\{c_m, c_d, n, q\}} \lambda(w_m/w_d)U_m + (1 - \lambda(w_m/w_d))U_d$$
$$\text{s.t. } c_m + c_d + p_q Q \leq I + w_d(1 - \ell_d) + w_m(1 - \ell_m - \gamma n)$$

where $\lambda'(\cdot) > 0$, so higher relative wages increase bargaining power. This model yields three key equations for sex selection, mother's leisure, and fertility:

$$\log q = \log\left(\frac{\lambda\alpha_{qm} + (1-\lambda)\alpha_{qd}}{\lambda\alpha_{nm} + (1-\lambda)\alpha_{nd}}\right) + \log\gamma - \log p_q + \log w_m$$
$$\log \ell_m = \log \alpha_\ell + \log\lambda - \log w_m + \log W$$
$$\log n = \log\left(\frac{\lambda\alpha_{nm} + (1-\lambda)\alpha_{nd}}{\gamma}\right) + \log W - \log w_m - \log\left(1 + \frac{\lambda\alpha_{qm} + (1-\lambda)\alpha_{qd}}{\lambda\alpha_{nm} + (1-\lambda)\alpha_{nd}}\right)$$

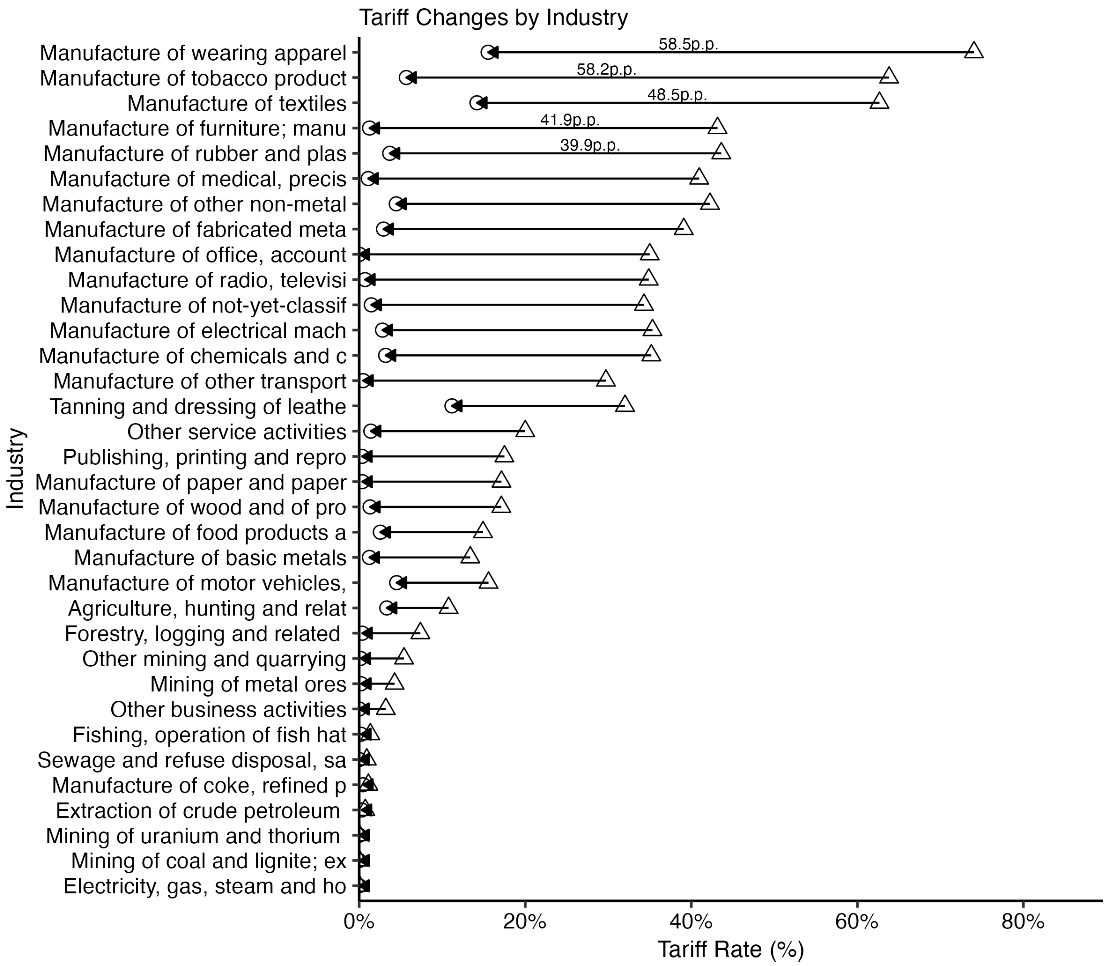
The comparative statics reveal that this model can match most but not all of our empirical findings. Specifically:

1. Mother's wages increase sex selection ($\partial q/\partial w_m > 0$) if mothers have stronger relative preference for sons: $\alpha_{qm}/\alpha_{nm} > \alpha_{qd}/\alpha_{nd}$
2. Mother's wages reduce fertility ($\partial n/\partial w_m < 0$) if mothers prefer fewer children than fathers: $\alpha_{nm} < \alpha_{nd}$
3. Father's wages increase fertility ($\partial n/\partial w_d > 0$) under the same conditions

However, under these conditions, the model predicts father's wages would reduce sex selection ($\partial q/\partial w_d < 0$), contrary to our finding of no effect. This result suggests that while bargaining may play a role, it cannot fully explain the asymmetric effects of parental industry exposures that we observe empirically.

C Additional Figures and Tables

Figure A1: Changes in U.S. Tariffs on Vietnamese Imports Following the 2001 BTA



Notes: This figure shows the changes in U.S. tariff rates on Vietnamese imports by industry following the 2001 U.S.-Vietnam Bilateral Trade Agreement. The start of each arrow represents the Column 2 tariff rate, while the arrowhead shows the new Most Favored Nation (MFN) tariff rate. Percentage points indicate the size of the tariff reduction for the five most affected industries. Data source: McCaig and Pavcnik (2018).

Table A4: Effects Robust to Regional Infrastructure Development: Region × Year

	Male Birth		Mom Work Hrs		Any Birth	
	(1)	(2)	(3)	(4)	(5)	(6)
$\tau_m \times$ Post	0.26	0.26	33.7	34.8	-0.01	-0.01
	(2.5)	(2.5)	(2.2)	(2.2)	(-3.4)	(-3.1)
$\tau_d \times$ Post	-0.04	-0.04	20.8	21.1	0.02	0.02
	(-0.41)	(-0.43)	(1.4)	(1.4)	(3.6)	(3.6)
$\tau_p \times$ Post	0.06	0.05	8.1	6.4	0.004	0.01
	(1.5)	(1.1)	(0.85)	(0.62)	(0.83)	(3.7)
$\tau_p^w \times$ Post	-0.06	-0.06	-8.0	-7.5	-0.003	-0.005
	(-1.7)	(-1.4)	(-0.84)	(-0.78)	(-0.60)	(-1.7)
R^2	0.01	0.01	0.34	0.35	0.28	0.28
Observations	28,390	28,389	26,129	26,128	1,911,566	1,911,566
Control Mean	0.58	0.58	157.9	157.9	0.01	0.01
Region-Year FE	✓	✓	✓	✓	✓	✓
Mom FE					✓	✓
Birth Year FE	✓	✓			✓	✓
Province, Industry FEs	✓	✓	✓	✓		
Controls		✓		✓		✓

Notes: Same as Table 2

Table A5: Effects Robust to Industry Clustering

	Male Birth		Mom Work Hrs		Any Birth	
	(1)	(2)	(3)	(4)	(5)	(6)
$\tau_m\times$ Post	0.26	0.26	34.9	36.6	-0.01	-0.01
	(3.2)	(3.3)	(1.3)	(1.5)	(-1.3)	(-1.2)
$\tau_d\times$ Post	-0.05	-0.05	18.9	19.6	0.02	0.02
	(-0.86)	(-0.90)	(1.3)	(1.5)	(4.8)	(4.7)
$\tau_p\times$ Post	0.03	0.02	-17.8	-20.4	0.004	0.01
	(1.2)	(0.73)	(-1.8)	(-2.1)	(2.1)	(6.4)
$\tau_p^w\times$ Post	-0.03	-0.03	22.2	20.4	-0.003	-0.005
	(-1.1)	(-0.85)	(1.9)	(2.0)	(-1.4)	(-2.7)
R^2	0.009	0.010	0.33	0.34	0.28	0.28
Observations	28,390	28,389	26,129	26,128	1,911,566	1,911,566
Control Mean	0.58	0.58	157.9	157.9	0.01	0.01
Mom FE					✓	✓
Birth Year FE	✓	✓			✓	✓
Survey Year FE			✓	✓		
Province, Industry FEs	✓	✓	✓	✓		
Controls		✓		✓		✓

Notes: Same as Table 2 except T-statistics based on standard errors clustered by mothers' industries in parentheses.

Table A6: Effects Robust to Sample Weights

	Male Birth		Mom Work Hrs		Any Birth	
	(1)	(2)	(3)	(4)	(5)	(6)
$\tau_m \times$ Post	0.28	0.28	30.1	32.7	-0.01	-0.01
	(3.0)	(3.0)	(1.9)	(2.1)	(-2.9)	(-2.7)
$\tau_d \times$ Post	-0.06	-0.06	23.9	24.3	0.02	0.02
	(-0.60)	(-0.60)	(1.7)	(1.8)	(3.7)	(3.7)
$\tau_p \times$ Post	0.02	0.01	-19.0	-22.5	0.006	0.02
	(0.45)	(0.36)	(-1.7)	(-2.3)	(1.2)	(3.9)
$\tau_p^w \times$ Post	-0.02	-0.02	24.9	22.6	-0.005	-0.007
	(-0.60)	(-0.62)	(2.2)	(2.3)	(-1.1)	(-2.1)
R^2	0.007	0.007	0.30	0.30	0.29	0.29
Observations	28,390	28,389	26,129	26,128	1,911,809	1,911,809
Control Mean	0.58	0.58	157.9	157.9	0.01	0.01
Mom FE					✓	✓
Birth Year FE	✓	✓			✓	✓
Survey Year FE			✓	✓		
Province, Industry FEs	✓	✓	✓	✓		
Controls		✓		✓		✓

Notes: Same as Table 2 except all regressions estimated without sampling weights.

Table A7: Effects Robust to Logistic Regression

	Male Birth	
	(1)	(2)
$\tau_m \times$ Post	1.1	1.1
	(2.5)	(2.4)
$\tau_d \times$ Post	-0.2	-0.2
	(-0.5)	(-0.5)
$\tau_p \times$ Post	0.1	0.07
	(0.8)	(0.5)
$\tau_p^w \times$ Post	-0.1	-0.1
	(-1.0)	(-0.8)
Observations	28,388	28,387
Birth Year FE	✓	✓
Province, Industry FEs	✓	✓
Controls		✓

Notes: Same as Table 2 except using Logit estimation instead of OLS.