

# Online Appendix for Habit Formation and the Misallocation of Labor

Matti Sarvimäki

Roope Uusitalo

Markus Jäntti

March 3, 2019

## Data details

**Census data** The starting point of our data is the 1950 population census. The original census forms were sorted by municipality, within municipalities in alphabetical order and then filed in folders. In 1995, Statistics Finland drew a sample from the full 1950 census by picking every tenth folder, see Statistics Finland (1998) for details. Almost all of the information on the census forms was manually inserted into a database. The resulting sample contains information on 411,629 persons from 392 of municipalities (out of a total of 547 municipalities) corresponding to 10.3% of the full population.

The 1950 census data is linked to other individual-level data sources using social security numbers. The social security numbers were introduced in September 1964 and thus had to be collected from the Population Register using names, date of birth, gender and place of birth. The match rate is very high. Social security numbers were found for 82.5% of the individuals included in the 1950 census sample. Furthermore, information from the 1970 census could be found for 73.1% of the original sample. In comparison, taking into account mortality and emigration, at most 74.5% of the population present in the 1950 census was also present in the 1970 census (Statistics Finland, 1998).

**Income data** Our information on individual-level income comes from the 1971 tax register. A key concern in using tax data in an analysis like ours is that taxable income might not be a comparable measure of true income for individuals working in and outside of agriculture. However, by 1971, agricultural profits were treated as taxable earned income and taxed according to the same rates as wage earnings. While production for own consumption was not taxed, agriculture had become increasingly specialized and, for example, Pihkala (1982) estimates that 90% of agricultural products were sold on the market and hence taxed. Much of the remaining 10% consisted of feeder crops used on the farm as intermediate inputs.

Figure A1 provides another check for the comparability of our income measure by plotting consumption expenditure against gross-income for farming and non-farming households using data from the 1971 Household Budget Survey (Statistics Finland, 1976). These data contain information on 1,186 households, of which we categorize 372 as farmers and 814 non-farmers based on the household reference person's main occupation. The consumption information was collected by the households during a four-week period and includes the purchase value of

items produced by the households (e.g. vegetables grown in the own garden). However, gifts and transfers to other households are not included as consumption. Gross income consists of all earnings and capital incomes as well as all public income transfers. The income information stems from the tax records of each household member and thus corresponds closely to the income measure used in our analysis. The expenditures have been annualized by Statistics Finland.

Panel A of Figure A1 represents the full data, while panel B focuses only on the inner 98% of each marginal distribution.<sup>1</sup> We find no indication of the tax records underestimating the consumption possibilities of farmers. In fact, the only statistically significant difference—at very high levels of income—suggest the opposite. However, these differences are driven by outliers. Once we drop the outliers, the confidence regions of the two groups overlap throughout (Panel B, Figure A1). Table A1 confirms this result by reporting regression coefficients using up to a fourth-order polynomial in income interacted with an indicator for the household’s reference person working in farming.

**Second-generation** We also have information on the children of persons included in the 1950 census sample. Statistics Finland has identified these children using their dataset on parent-child links. For each child, we observe education taken from the register of degrees, income from tax registers and basic demographic variables from the census and administrative registers.

A limitation of these data is that we observe only one parent for 42% of the children. This feature of the data is likely driven by the fact that in order for both parents to be observed, they have to be included in the 1950 census sample. Given the sampling scheme described above, we observe both (known) parents if they lived in the same household. Furthermore, we are more likely to observe them if they were living in the same municipality in 1950.<sup>2</sup> As a consequence, children for whom we observe both parents are likely to differ from children for whom we observe only one parent. This selection process is hard to characterize and may have been influenced by the resettlement. For these reasons, we have opted for an analysis sample in which the first generation is defined using information on only one parent. While this approach facilitates the interpretation of the estimates, it also means that some children categorized into the control group do in fact have one displaced parent.

## Additional empirical results

**Descriptive statistics** Tables A2–A5 show a detailed examination of the pre-war differences between displaced and non-displaced rural population. For comparison, columns 1–2 report the same numbers we reported in Table 1, i.e. sample averages of the background variables by future displacement status. The remaining columns show estimates and standard

---

<sup>1</sup>That is, panel B uses “shaved” data, where we have excluded 42 observations with income or expenditure that is lower than the 1st or higher than the 99th percentile of each distribution.

<sup>2</sup>Moreover, this likelihood is affected by the size of the municipality. Specifically, if all census forms of a municipality would fit into one folder, we would observe everyone living in the municipality. As the size of the municipality grows, i.e. the census forms fill more folders, the expected share of individuals ending into our sample approaches 10%.

errors from regressions of the form

$$x_{0i} = \alpha + \beta D_i + \mathbf{Z}_{0i}\gamma + \varepsilon_{it} \quad (\text{A1})$$

where  $x_{0i}$  is a pre-war characteristic of individual  $i$ ,  $D_i$  is an indicator for future displacement status (i.e. living in the ceded area before the war) and  $\mathbf{Z}_{0i}$  is a vector of other pre-war characteristics we may want to condition on. Columns 3–4 report the baseline differences, columns 5–6 control for the distance of the 1939 residence municipality to the post-war border (and its interaction with future displacement status), and columns 7–8 condition on resettlement area fixed-effects.

**Impact on 1950 outcomes** Figure A2 and the top panel of Figure A4 present estimates for the impact of forced migration on short-term outcomes as measured in the 1950 census using the same seven specifications used for income (see Section 3 and note for Table 2 of the main papers for details). For reference, Figure A3 and the bottom panel of Figure A4 present the estimates in the same format for the corresponding outcomes in 1970. These figures also serve as robustness checks for the main estimates for these outcomes (i.e. we report the second and the last estimates of each panel of these figures in Table 3 of the main paper).

## An illustrative model of habit formation and location choice

**Environment** We assume that agents’ contemporaneous utility is an additive function of location capital and consumption

$$u_{jt}(c_t, l_{jt}) = c_t + l_{jt}^\alpha \quad (\text{A2})$$

where  $t$  and  $j$  index time and location,  $c$  is consumption,  $l$  is the time the person has lived in the location (“location capital”), and  $\alpha \geq 0$  is a parameter governing the strength of habit formation.

Individuals live in one location during childhood and then work for  $T$  periods. During their working life, they choose a sequence of locations,  $I_{jt}$ , to maximize life-time utility

$$\max_{\{I_{jt}\}} U = \sum_{t=1}^T u_j(c_t, l_{jt}) \quad (\text{A3})$$

subject to a budget constraint and accumulation of location capital. The budget constraint is

$$\sum_{t=1}^T c_t \leq \sum_{t=1}^T z_j \quad (\text{A4})$$

where  $z_j$  is the income the agent would earn each period in location  $j$ . These incomes are constant over time and drawn at birth from a joint distribution of sectoral incomes  $G(z_s)$ . Locational capital is accumulated as

$$l_{jt} = l_{j,t-1} + I_{jt}, \quad (\text{A5})$$

where  $I_{jt}$  is an indicator function taking the value one if the agent lives in location  $j$  in period  $t$  and zero otherwise.

We define period one as the stage when the person starts to make her own decisions and assume that she enters this stage with initial location capital,  $l_{j0}$ , accumulated during her childhood and thus reflecting the decisions of her parents. In order to keep the model as simple as possible, we treat locations and jobs as isomorphic and abstract away from discounting, depreciation of location capital, price and wage dynamics, local amenities, differences in regional prices and other migration costs.

Given these assumptions, utility is maximized by spreading consumption evenly over the life-cycle. Furthermore, if the agent migrates, she does so immediately at  $t = 1$  in order to start accumulating location capital in the new location as soon as possible. For the same reason, it is never optimal to migrate twice. Thus the maximum utility the agent can derive from choosing location  $j$  for the remaining of her life is

$$V(z_j, l_{j0}, \alpha, T) = Tz_j + \sum_{t=1}^T (l_{j0} + t)^\alpha \quad (\text{A6})$$

**Migration choice** Consider now the choice of whether to stay at home location  $h$  or to move to some other location. We denote the location providing the highest income with  $m$ . If she migrates, she will choose  $m$  as she has no reason to move to lower paying location for which she does not have any location capital. She migrates to  $m$  if  $V(z_m, 0, \alpha, T) > V(z_h, l_{h0}, \alpha, T)$  or

$$z_m - z_h \geq \frac{\sum_{t=1}^T (l_{h0} + t)^\alpha - \sum_{t=1}^T t^\alpha}{T} \quad (\text{A7})$$

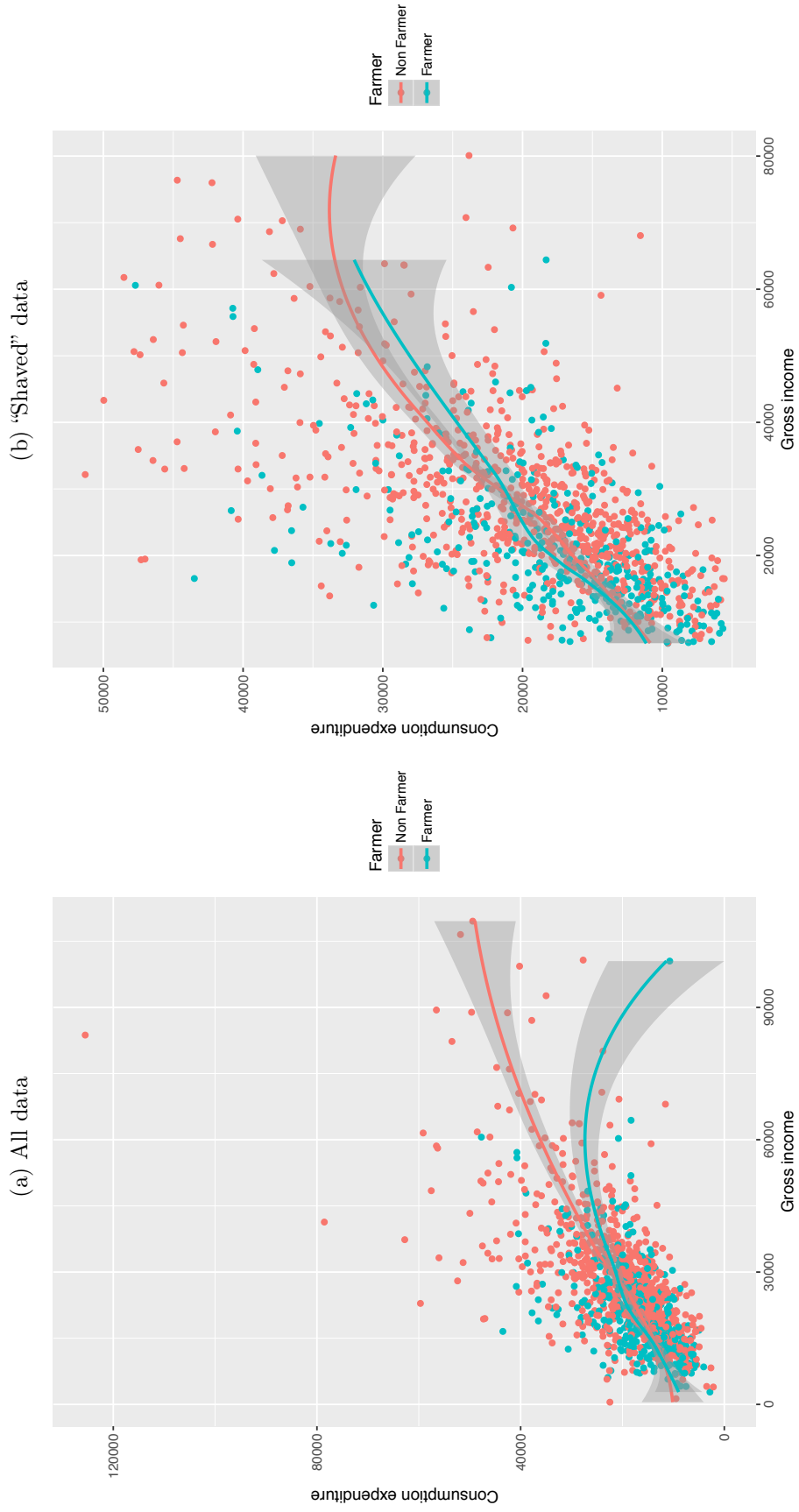
Condition (A7) illustrates that even in a highly stylized model, individuals may choose between migrating or staying for many reasons. Some stay because their skills have the highest return in their home location ( $z_m - z_h < 0$ ). Others could increase their income by migrating, but would lose too much utility by giving up their initial location capital. This trade-off gives rise to the income difference required for migration that is larger for individuals who have lived longer in the same place during their childhood (and thus have higher  $l_{h0}$ ) and for those who have stronger location preferences (higher  $\alpha$ ). Furthermore, the minimum income difference required for migration decreases with the length of the future career,  $T$ .

## References

- Pihkala, E. (1982). Maa-ja metsätalouden uusi asema. In J. Ahvenainen, P. Erkki, and R. Viljo (Eds.), *Suomen taloushistoria 2: Teollistuva Suomi*. Helsinki: Tammi.
- Statistics Finland (1976). Kotitaloustiedustelu 1971, osa 1, kotitalouksien kulutusmenot. Tilastollisia tiedonantoja 55. Helsinki, Finland.
- Statistics Finland (1998). The Finnish Longitudinal Census Data File. Statistics Finland.

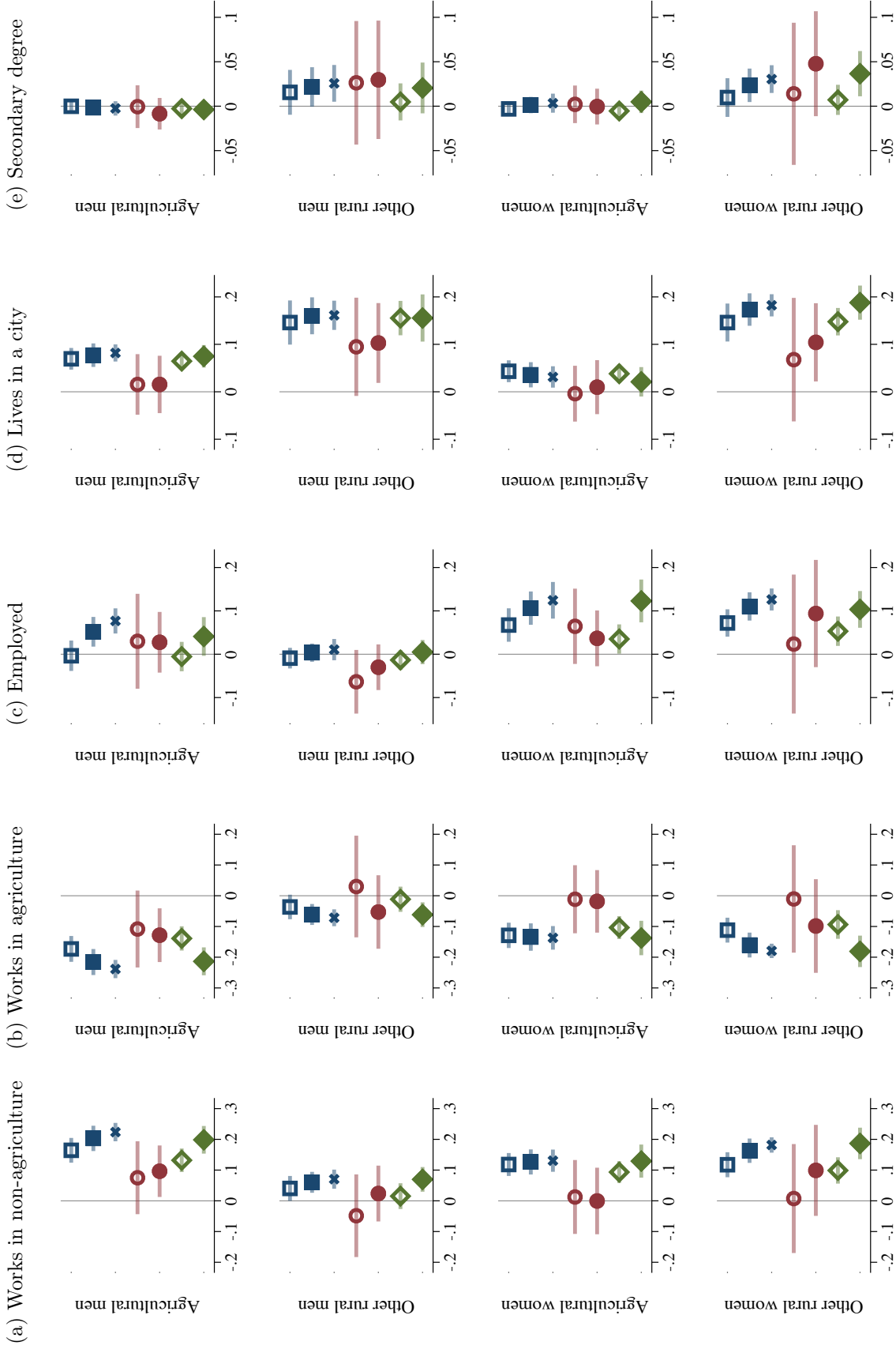
Waris, H., V. Jyrkilä, K. Raitasuo, and J. Siipi (1952). *Siirtoväen sopeutuminen*. Otava.

Figure A1: Consumption expenditure and gross income among farm and non-farm households in 1971



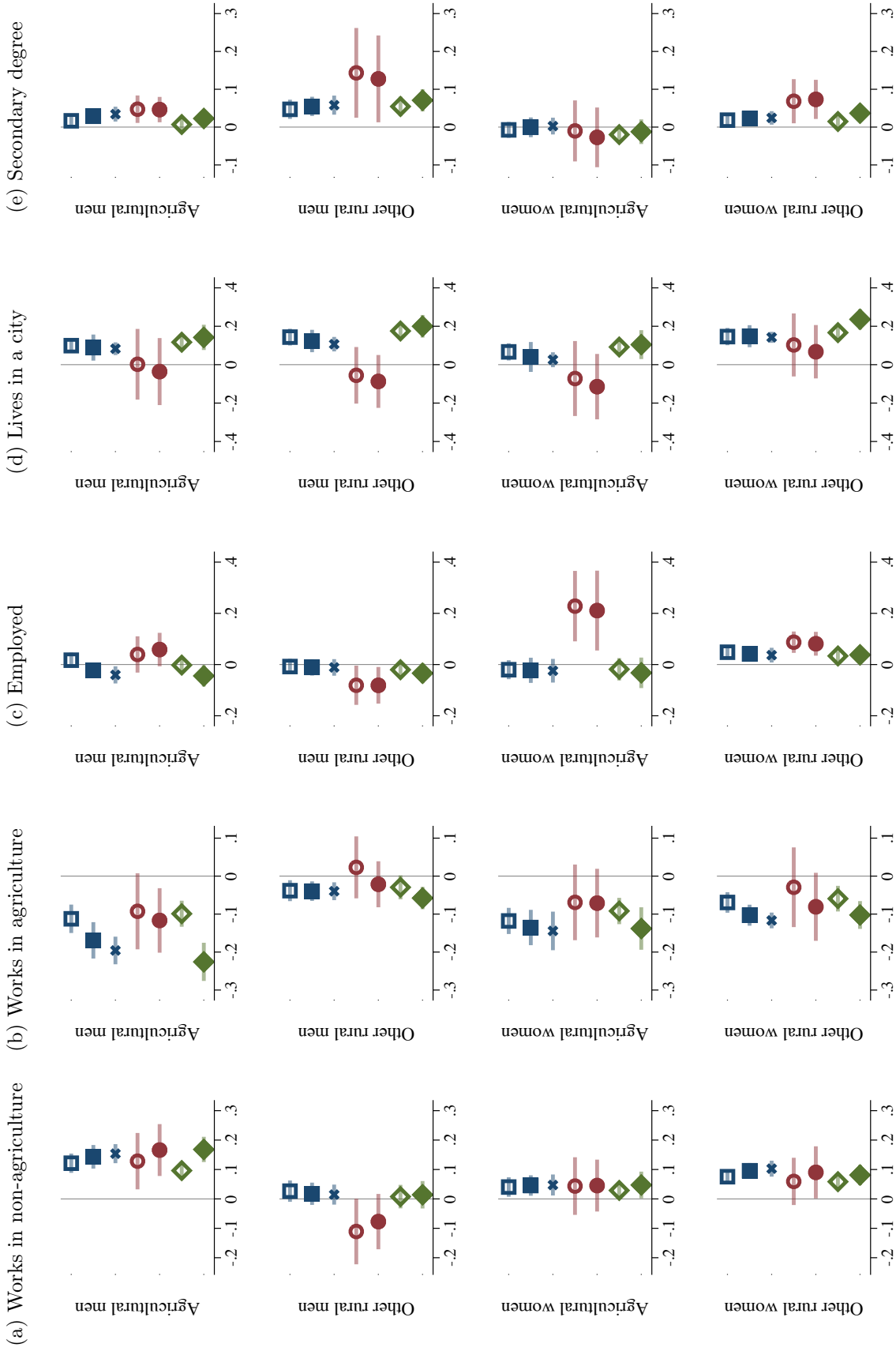
Note: Each point corresponds to a household. The lines represent loess-lines within the group (farmer, non-farmer). The shaded areas show the corresponding 95% confidence regions.

Figure A2: Impact of forced migration on 1950 outcomes, rural population



Note: Point estimates and 95% confidence intervals for the impact of forced migration on 1950 outcomes using specifications corresponding to those in Table 2 of the main paper. Solid markers refer to estimates controlling for pre-war observable characteristics. Squares show estimates from the baseline specification, circles are the Oster Bounds, crosses are the spatial RD specification and diamonds for specifications controlling for resettlement area fixed-effects. See Section 3 and notes for Table 2 of the main paper for details.

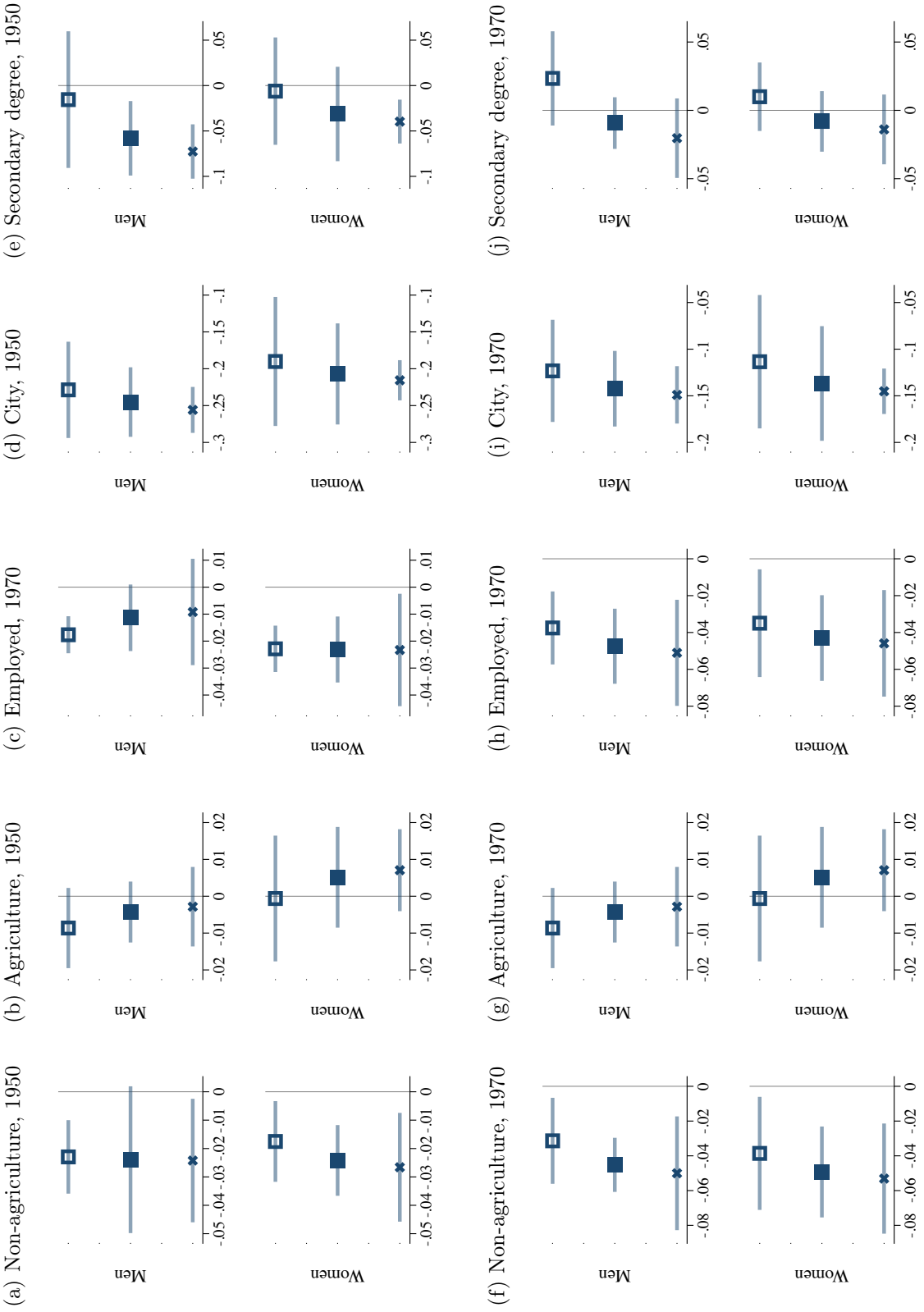
Figure A3: Impact of forced migration on 1970 outcomes, rural population



Note: Point estimates and 95% confidence intervals for the impact of forced migration on 1970 outcomes using specifications corresponding to those in Table 2 of the main paper. Solid markers refer to estimates controlling for pre-war observable characteristics. Squares show estimates from the baseline specification, crosses are the Oster Bounds, circles are the spatial RD specification and diamonds for specifications controlling for resettlement area fixed-effects. See Section 3 and notes for Table 2 of the main paper for details.

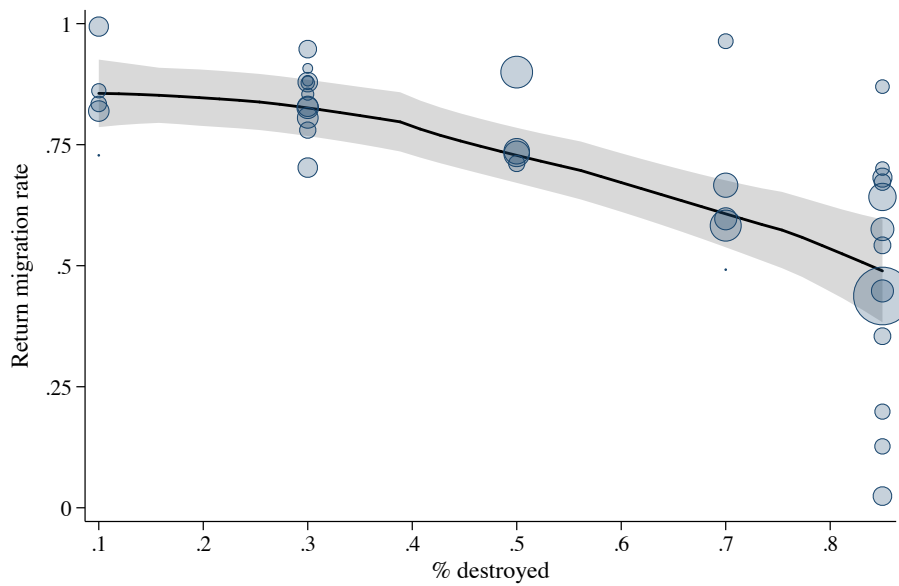


Figure A4: Impact of forced migration on 1950 and 1970 outcomes, urban population



Note: Point estimates and 95% confidence intervals for the impact of forced migration using specifications corresponding to those in Table 2 of the main paper. Solid markers refer to estimates controlling for pre-war observable characteristics. Squares show estimates from the baseline specification, crosses are the Oster Bounds. See Section 3 and notes for Table 2 of the main paper for details.

Figure A5: Return Migration and the Destruction of the Housing Stock



Y-axis: Share of the pre-war population who had returned by January 1st, 1944. Drafted men are included in the denominator, but not in the numerator. X-axis: The share of existing housing stock destroyed by December 31st, 1941. Source: Waris et al. (1952, Appendix Tables 7 and 9)

Table A1: Consumption expenditure and gross income

	(1)	(2)	(3)	(4)
Constant	8.0518 (0.5795)	4.3229 (1.1665)	8.5764 (2.2106)	11.6125 (3.9673)
Farmer	0.8223 (0.9718)	2.4083 (1.9341)	-2.3713 (3.6753)	-10.1319 (6.9078)
Farmer $\times$ gross income	-0.0291 (0.0382)	-0.0922 (0.1421)	0.4298 (0.4134)	1.7449 (1.0932)
Farmer $\times$ gross income <sup>2</sup> /10		0.0020 (0.0234)	-0.1610 (0.1403)	-0.8899 (0.5929)
Farmer $\times$ gross income <sup>3</sup> /100			0.0147 (0.0143)	0.1736 (0.1300)
Farmer $\times$ gross income <sup>4</sup> /1000				-0.0117 (0.0098)
Gross income	0.4385 (0.0189)	0.6917 (0.0714)	0.2398 (0.2120)	-0.2068 (0.5290)
Gross income <sup>2</sup> /10		-0.0361 (0.0098)	0.1007 (0.0612)	0.3138 (0.2392)
Gross income <sup>3</sup> /100			-0.0119 (0.0053)	-0.0518 (0.0436)
Gross income <sup>4</sup> /1000				0.0025 (0.0027)
Observations	1,144	1,144	1,144	1,144
Adjusted $R^2$	0.3913	0.3987	0.4004	0.4003

Note: Coefficients and standard errors (in parantheses) from regressing consumption expenditure on gross income (both measured in 1000s of *markka*) using data from the 1971 Household Budget Survey.

Table A2: Pre-War Characteristics of the Rural Population: Agricultural men

	Differences							
	Means		Baseline		Spatial RD		Within Resettlement Area	
	Non-disp. (1)	Disp. (2)	Diff. (3)	se. (4)	Diff (5)	se. (6)	Diff (7)	se. (8)
<i>A: Demographics</i>								
Age	22.8	22.8	-0.01	(0.14)	0.24	(0.50)	-0.15	(0.19)
Swedish-speaker	0.07	0.00	-0.07	(0.01)	0.01	(0.01)	-0.02	(0.01)
Migrated prior to 1939	0.18	0.15	-0.02	(0.02)	0.02	(0.13)	-0.08	(0.03)
<i>B: Socioeconomic status</i>								
Entrepreneur	0.30	0.42	0.12	(0.02)	0.14	(0.07)	0.14	(0.02)
White-collar	0.02	0.02	-0.01	(0.00)	-0.01	(0.02)	0.00	(0.01)
Blue-collar	0.36	0.16	-0.20	(0.02)	-0.04	(0.06)	-0.25	(0.02)
Assisting family member	0.32	0.41	0.09	(0.02)	-0.04	(0.07)	0.11	(0.02)
<i>C: Characteristics of the municipality of residence</i>								
Average taxable income	1.41	1.38	-0.02	(0.12)	-0.91	(1.14)	-0.56	(0.13)
Agricultural LFS	0.83	0.81	0.01	(0.01)	-0.03	(0.06)	-0.01	(0.01)
Latitude	69.4	67.7	-1.63	(0.15)	-0.47	(0.82)	0.00	(0.00)

Note: Pre-war differences between displaced and non-displaced rural population. Columns 1–2 report averages, columns 3–4 baseline differences, columns 5–6 control for the distance of the 1939 residence municipality to the post-war border (and its interaction with future displacement status), and columns 7–8 condition on resettlement area fixed-effects.

Table A3: Pre-War Characteristics of the Rural Population: Other rural men

	Means		Differences					
	Disp. (1)	Non- disp. (2)	Baseline		Spatial RD		Within Resettlement Area	
			Diff. (3)	se. (4)	Diff (5)	se. (6)	Diff (7)	se. (8)
<i>A: Demographics</i>								
Age	22.9	22.8	-0.16	(0.16)	-0.97	(0.53)	-0.21	(0.19)
Swedish-speaker	0.08	0.00	-0.08	(0.02)	0.01	(0.01)	-0.07	(0.03)
Migrated prior to 1939	0.40	0.36	-0.04	(0.03)	-0.28	(0.15)	-0.06	(0.03)
<i>C: Socioeconomic status</i>								
Entrepreneur	0.07	0.06	-0.01	(0.01)	0.02	(0.02)	-0.01	(0.01)
White-collar	0.10	0.13	0.03	(0.01)	0.09	(0.04)	0.03	(0.01)
Blue-collar	0.56	0.53	-0.03	(0.02)	-0.29	(0.09)	-0.09	(0.03)
Assisting family member	0.27	0.28	0.01	(0.02)	0.10	(0.07)	0.07	(0.02)
<i>D: Sector of employment</i>								
Manufacturing	0.27	0.20	-0.08	(0.02)	-0.31	(0.09)	-0.13	(0.02)
Construction	0.14	0.14	-0.01	(0.01)	0.01	(0.03)	-0.02	(0.01)
Services	0.24	0.32	0.08	(0.02)	0.22	(0.06)	0.09	(0.02)
<i>E: Characteristics of the municipality of residence</i>								
Average taxable income	1.95	1.65	-0.30	(0.18)	-1.32	(0.58)	-0.88	(0.17)
Agricultural LFS	0.74	0.76	0.03	(0.03)	0.03	(0.12)	0.09	(0.03)
Latitude	68.9	67.7	-1.18	(0.17)	-0.38	(0.73)	0.00	(0.00)

Note: Pre-war differences between displaced and non-displaced rural population. Columns 1–2 report averages, columns 3–4 baseline differences, columns 5–6 control for the distance of the 1939 residence municipality to the post-war border (and its interaction with future displacement status), and columns 7–8 condition on resettlement area fixed-effects.

Table A4: Pre-War Characteristics of the Rural Population: Agricultural women

	Differences							
	Means		Baseline		Spatial RD		Within Resettlement Area	
	Non-disp. (1)	Disp. (2)	Diff. (3)	se. (4)	Diff (5)	se. (6)	Diff (7)	se. (8)
<i>A: Demographics</i>								
Age	24.2	24.7	0.54	(0.19)	-0.41	(0.50)	0.38	(0.22)
Swedish-speaker	0.07	0.00	-0.07	(0.02)	0.01	(0.01)	-0.01	(0.01)
Migrated prior to 1939	0.29	0.24	-0.05	(0.03)	-0.04	(0.12)	-0.10	(0.03)
<i>B: Socioeconomic status</i>								
Entrepreneur	0.04	0.03	-0.01	(0.01)	0.01	(0.02)	0.00	(0.01)
White-collar	0.00	0.00	0.00	(0.00)	0.00	(0.00)	0.00	(0.00)
Blue-collar	0.23	0.13	-0.10	(0.02)	0.08	(0.05)	-0.19	(0.03)
Assisting family member	0.73	0.84	0.11	(0.02)	-0.09	(0.07)	0.19	(0.03)
<i>C: Characteristics of the municipality of residence</i>								
Average taxable income	1.45	1.38	-0.07	(0.11)	-1.03	(0.86)	-0.57	(0.13)
Agricultural LFS	0.83	0.80	0.01	(0.01)	-0.06	(0.04)	-0.01	(0.01)
Latitude	69.3	67.7	-1.51	(0.15)	0.27	(0.49)	0.00	(0.00)

Note: Pre-war differences between displaced and non-displaced rural population. Columns 1–2 report averages, columns 3–4 baseline differences, columns 5–6 control for the distance of the 1939 residence municipality to the post-war border (and its interaction with future displacement status), and columns 7–8 condition on resettlement area fixed-effects.

Table A5: Pre-War Characteristics of the Rural Population: Other rural women

	Means		Differences					
			Baseline		Spatial RD		Within Resettlement Area	
	Disp. (1)	Non-disp. (2)	Diff. (3)	se. (4)	Diff (5)	se. (6)	Diff (7)	se. (8)
<i>A: Demographics</i>								
Age	22.7	22.5	-0.23	(0.16)	-0.50	(0.91)	-0.63	(0.18)
Swedish-speaker	0.07	0.00	-0.07	(0.01)	0.00	(0.00)	-0.05	(0.02)
Migrated prior to 1939	0.41	0.35	-0.06	(0.03)	-0.18	(0.15)	-0.10	(0.03)
<i>C: Socioeconomic status</i>								
Entrepreneur	0.02	0.02	0.00	(0.00)	-0.01	(0.01)	0.00	(0.00)
White-collar	0.10	0.10	0.00	(0.01)	-0.02	(0.03)	0.00	(0.01)
Blue-collar	0.16	0.13	-0.03	(0.01)	0.00	(0.03)	-0.07	(0.02)
Assisting family member	0.72	0.75	0.03	(0.02)	0.04	(0.04)	0.08	(0.02)
<i>D: Sector of employment</i>								
Manufacturing	0.08	0.06	-0.02	(0.01)	-0.06	(0.03)	-0.05	(0.02)
Construction	0.00	0.00	0.00	(0.00)	0.00	(0.00)	0.00	(0.00)
Services	0.18	0.18	-0.01	(0.01)	0.04	(0.03)	-0.03	(0.01)
<i>E: Characteristics of the municipality of residence</i>								
Average taxable income	1.81	1.58	-0.23	(0.16)	-1.25	(0.78)	-0.85	(0.16)
Agricultural LFS	0.76	0.78	0.02	(0.03)	0.01	(0.10)	0.09	(0.03)
Latitude	69.0	67.8	-1.24	(0.16)	-0.49	(0.74)	0.00	(0.00)

Note: Pre-war differences between displaced and non-displaced rural population. Columns 1–2 report averages, columns 3–4 baseline differences, columns 5–6 control for the distance of the 1939 residence municipality to the post-war border (and its interaction with future displacement status), and columns 7–8 condition on resettlement area fixed-effects.

Table A6: Industry Mix in Non-Agriculture, 1970

	Agricultural		Other rural		Urban	
	Dis- placed (1)	Non- disp. (2)	Dis- placed (3)	Non- disp. (4)	Dis- placed (5)	Non- disp. (6)
			<i>A: Men</i>			
Mining and quarrying	0.01	0.01	0.00	0.01	0.00	0.00
Manufacturing	0.31	0.33	0.28	0.34	0.28	0.34
Electricity, gas and water	0.02	0.02	0.03	0.02	0.02	0.02
Construction	0.31	0.28	0.21	0.20	0.14	0.12
Trade, restaurants, hotels	0.08	0.07	0.10	0.09	0.17	0.13
Transport, storage and communication	0.11	0.10	0.16	0.13	0.11	0.12
Community, social and personal services	0.03	0.04	0.04	0.03	0.06	0.04
			<i>B: Women</i>			
Mining and quarrying	0.00	0.00	0.00	0.00	0.00	0.00
Manufacturing	0.44	0.31	0.31	0.28	0.22	0.29
Electricity, gas and water	0.00	0.01	0.01	0.01	0.01	0.01
Construction	0.03	0.03	0.01	0.02	0.02	0.01
Trade, restaurants, hotels	0.20	0.20	0.24	0.26	0.28	0.29
Transport, storage and communication	0.00	0.05	0.05	0.06	0.06	0.05
Community, social and personal services	0.02	0.03	0.03	0.03	0.03	0.05

Note: 1-digit industry shares among those working outside of agriculture in 1970.