Supplement to "The long run health consequences of rural-urban migration"

(*Quantitative Economics*, Vol. 10, No. 2, May 2019, 565–606)

Janna E. Johnson
Humphrey School of Public Affairs, University of Minnesota

Evan J. Taylor
Center for the Economics of Human Development, University of Chicago

Appendix A: A simple model of information and migration

The key to our model is an assumption that individuals initially reside in two types of rural town—those with *high* access to information about the outside world, and those with *low* access—which are otherwise identical. In high information towns, information about wage prospects in a migration destination are more "precise" (a notion that we clarify shortly).

Individuals born in these town are endowed with latent ability α_i, drawn from the same distribution in each type of town. Indirect utility for individual i is given by $u(w\alpha_i)$, where w is the rental rate on ability, and the wage is given by $w\alpha_i$.

We assume $\frac{\partial u}{\partial w} > 0$, of course, and we assume risk aversion, $\frac{\partial^2 u}{\partial w^2} < 0$. We also assume that the utility return to ability increases with the wage, $\frac{\partial^2 u}{\partial \alpha \partial w} > 0$.[1]

If individual i migrates to an urban location, utility is $u(w_M\alpha_i) - C$, where w_M is the wage in the migration destination, and C is a known fixed cost that encompasses net monetary and psychic costs of migration. If instead the individual remains in the rural town, utility is $u(w_S\alpha_i)$, where w_S is the wage earned by a "stayer."[2]

Individuals forms beliefs about the wage in each location. They migrate if, given their beliefs, they have a higher expected utility from migrating than from staying home:

$$E_i\big[u(w_M\alpha_i)\big] - C > E_i\big[u(w_S\alpha_i)\big]. \tag{A1}$$

Janna E. Johnson: jannaj@umn.edu
Evan J. Taylor: evanjtaylor@gmail.com

[1]For some utility functions, this may not necessarily be the case. For example, if $u(\ )$ is a CRRA utility function, then $\frac{\partial^2 u}{\partial \alpha \partial w} = (1-\sigma)(w\alpha)^{-\sigma}$. For $\sigma > 1$, this would imply a negative cross-partial, and negative selection into migration. For our model to produce positive selection, we only need a utility function which produces higher returns to ability in the urban areas, which could be caused by amenities as well as wage.

[2]We do not incorporate "return migration" in our model (see, e.g., Dustmann and Görlach (2016)), because it does not appear to be a major feature of migration in our context. Also, the cost C is borne early in life while the returns accrue over the lifetime; we implicitly assume potential migrants are not credit constrained (as in Dustmann and Okatenko (2014)), which seems reasonable given the relative prosperity of the population we study.

© 2019 The Authors. Licensed under the Creative Commons Attribution-NonCommercial License 4.0. Available at http://qeconomics.org. https://doi.org/10.3982/QE962

We assume the following about beliefs:

(1) In both types of towns, all individuals have the same probabilistic beliefs about w_S, denoted by the distribution B_S.

(2) In towns with *high* information access, all individuals have the same beliefs about w_M, denoted by $B_{M,H}$; while individuals born in towns with *low* access have the same beliefs about w_M, denoted $B_{M,L}$.

(3) The distributions $B_{M,H}$ and $B_{M,L}$ both first order stochastic dominate B_S; individuals expect to earn higher wages if they migrate.

(4) $B_{M,L}$ is a mean-preserving spread of $B_{M,H}$.

Assumption 3 ensures that for each town, there will exist some cut-off level of ability, say $\hat{\alpha}$, such that all individuals with $\alpha_i > \hat{\alpha}$ migrate, while all individuals with $\alpha_i < \hat{\alpha}$ do not. This result arises from the complementarity of wages and ability. Intuitively, the expected wage increase for a low ability individual does not offset the fixed cost, so she does not migrate; the converse is true for a high ability individual. Our model thereby predicts positive selection into migration.

Assumption 4 implies that individuals in low information towns have beliefs that are more uncertain. Because of risk aversion, expected utility from migration for these individuals will be lower than for comparable individuals born in high information towns. Hence the migration threshold in high information towns, say $\hat{\alpha}_H$, will be lower than in low information towns, $\hat{\alpha}_H < \hat{\alpha}_L$. High information towns will thus have more out-migration.

To summarize, our model has three key features that are broadly consistent with the empirical facts.[3] First, migrants are positively selected (on latent ability). Second, high information towns will have a higher migrant share than low information towns. Third, the "marginal" migrants—those who would migrate if and only if they were born in a high information town—are of lower ability than individuals who migrate regardless of their town's information access.

Appendix B: Balancing tests of the mail category instrument

We investigate the correlation of our mail category instrument with place-specific characteristics from U.S. Census data. We consider several county characteristics that could plausibly be correlated with mail flow, using data from 1930. Summary statistics are in Table B.1. These include mean household size, fraction of households that are farm households, fraction of households that own a radio, and the fraction of boys and girls in the county aged 6–18 attending school.

The smallest geographic area identified in the census that we can match to the Duke SSA/Medicare data is the county, and as our instrument is town-level, we create a county-level mail category measure as the mean mail category in the Duke sample. While our town-level instrument has six categories, the county mean mail category

[3] There may be other models which are also consistent with these facts, for example, one where information directly reduces the cost of migration by giving migrants knowledge of the cheapest transportation routes.

TABLE B.1. County-level summary statistics, North Dakota, South Dakota, and Montana.

	Mean	SD	Minimum	Maximum
Mean mail category	2.257	0.882	0	4
Mean household size	4.235	0.586	2.176	5.743
Fraction farm households	0.573	0.192	0.029	1.000
Fraction of households owning radio	0.393	0.104	0.059	0.623
Fraction of boys ages 6–18 in school	0.912	0.055	0.696	1.000
Fraction of girls ages 6–18 in school	0.921	0.055	0.698	1.000
Population	10,276	7890	468	49,936

Note: $N = 174$. Mean mail category from Duke SSA/Medicare sample. Town mail category ranges from 0 to 5. Other county characteristic data from 1930 census. Sample includes all white individuals residing in counties in North Dakota, South Dakota, and Montana, excluding Shannon, SD, which had no girls aged 6–18 in 1930 census data.
Source: Authors' calculations using Duke SSA/Medicare data, mail-car capacity information from Borchert (1987), and 1930 U.S. Census data from Ruggles et al. (2010).

ranges from 0 (there are no railroads in the county) to 4 (all railroads in the county are in the 600–720 linear feet per week category). The mean value of this instrument is 2.26, between 90–120 and 270–510 linear feet per week.

We regress county average mail category on these characteristics, including a cubic in county population and state fixed effects. Results are in Table B.2. County average mail category is strongly correlated with three characteristics: mean household size, fraction of households that are farms, and fraction of households that own a radio. Areas of higher mail flow tend to have larger households, fewer farm households, and more households with radios.

TABLE B.2. Regression of mail category on county characteristics, North Dakota, South Dakota, and Montana.

	Coefficients
Mean household size	0.512
	(0.131)
Fraction farm households	−2.645
	(0.501)
Fraction of households owning radio	1.888
	(0.596)
Fraction of boys ages 6–18 in school	−1.481
	(1.147)
Fraction of girls ages 6–18 in school	−0.626
	(1.406)
R^2	0.369
N	174

Note: Robust standard errors in parentheses. Dependent variable is county mean mail category in the Duke SSA/Medicare sample. Data on independent variables from 1930 census. Sample includes white individuals only. Shannon, SD is excluded as 1930 census reported no girls aged 6–18 residing in county. Specification also includes state fixed effects and a cubic in 1930 county population.
Source: Authors' calculations using Duke SSA/Medicare data, mail-car capacity information from Borchert (1987), and 1930 U.S. Census data from Ruggles et al. (2010).

Appendix C: MTE estimation using the separate approach

Identification of the MTE using the parametric and semiparametric approaches rely on the standard IV assumptions of Imbens and Angrist (1994), including that of conditional independence: $Z \perp\!\!\!\perp (U_0, U_1, V)|X$. It is common to add the following additional assumptions, which are unnecessary in the ideal case, where the instrument Z generates full common support of the propensity score over the unit interval conditional on $X = x$. These include:

(1) *Linearity in X*. Condition on X parametrically by entering it in the outcome and propensity score equations as $X\beta$.

(2) *Additive separability*. The unobserved and observed components of the expected outcome equations are additively separable.

The above two assumptions, in addition to the IV assumptions, lead to the following potential outcome equations:

$$\begin{aligned} E(Y_0|X = x, U_D = u_D) &= x\beta_0 + E(U_0|U_D = u_D), \\ E(Y_1|X = x, U_D = u_D) &= x\beta_1 + E(U_1|U_D = u_D). \end{aligned} \tag{C1}$$

The MTE can then be estimated using an instrument that does not generate full common support over the unit interval. Estimation is even possible with a binary instrument, as shown by Brinch, Mogstad, and Wiswall (2017). Below we outline the steps required to estimate the MTE using the parametric and semiparametric variants of the separate approach.

C.1 *Parametric*

(1) Estimate $P(X, Z)$ by recovering the predicted values of a logistic regression of the treatment indicator on the observable characteristics X and instrument Z. Call these propensity scores $\hat{p}$.

(2) Regress Y_0 and Y_1 on X and a polynomial in $\hat{p}$ using OLS separately for untreated and treated observations. The same order polynomial should be used in both regressions. Multiple estimations using different orders of this polynomial can serve as robustness checks, or the optimal order can be chosen using an optimization procedure such as k-fold validation or leave-one-out cross validation.

(3) Using the following expressions, calculate $k_1(p)$ and $k_0(p)$ for each value of $\hat{p}$:

$$k_1(p) = p\frac{\partial K_1(p)}{\partial p} + K_1(p), \tag{C2}$$

$$k_0(p) = -(1-p)\frac{\partial K_0(p)}{\partial p} + K_0(p). \tag{C3}$$

The polynomials in $\hat{p}$ from step 2 for treated and untreated observations form the estimates of $K_1(p)$ and $K_0(p)$, respectively. As both are polynomials, the derivatives of these expressions with respect to p are simple to compute.

(4) Using estimates of β_0 and β_1 from step 2 and $k_0(p)$ and $k_1(p)$ from step 3, calculate $\widehat{\text{MTE}}$ as follows for each value of $(x, \hat{p})$:

$$\widehat{\text{MTE}}(X = x, U_D = p) = x\hat{\beta}_1 - x\hat{\beta}_0 + \widehat{k_1(p)} - \widehat{k_0(p)}. \tag{C4}$$

(5) If desired, estimate standard errors for $\widehat{\text{MTE}}$ using a bootstrap procedure. In our estimation, we use 100 replications.

C.2 *Semiparametric*

(1) Estimate $P(X, Z)$ by recovering the predicted values of a logistic regression of the treatment indicator on observable characteristics X and instrument Z. Call these propensity scores $\hat{p}$.

(2) Demean the observables X as follows:

$$\tilde{X} = X - \bar{X}. \tag{C5}$$

Using $\tilde{X}$ ensures the MTE is calculated at the mean of X.

(3) Regress Y and $\tilde{X}$ on the propensity score $\hat{p}$ using a local polynomial regression and recover the residuals from each regression. Denote these Y^* and $\tilde{X}^*$, respectively. Fan and Gijbels (1996) recommended using a polynomial degree one degree higher than the polynomial you are trying to estimate. As we need to recover the first derivative of the polynomial, we estimate using a local quadratic.

(4) Divide observations into untreated (Y_0) and treated (Y_1) groups. Regress Y_0^* and Y_1^* on $\tilde{X}^*$ using OLS,

$$\begin{aligned} Y_0^* &= \tilde{X}^*\beta_0 + \varepsilon_0, \\ Y_1^* &= \tilde{X}^*\beta_1 + \varepsilon_1, \end{aligned} \tag{C6}$$

and recover $\hat{\beta}_0$ and $\hat{\beta}_1$.

(5) Calculate $Y_0^{**} = Y_0 - \hat{\beta}_0\tilde{X}$ for untreated observations and $Y_1^{**} = Y_1 - \hat{\beta}_1\tilde{X}$ for treated observations. Regress these separately on $\hat{p}$ using local quadratic regression to identify K_0, K_1, $\frac{\partial K_0}{\partial p}$, and $\frac{\partial K_1}{\partial p}$.

(6) Using the following expressions, calculate $k_1(p)$ and $k_0(p)$ for each value of $\hat{p}$ using the estimates from step 5:

$$k_1(p) = p\frac{\partial K_1(p)}{\partial p} + K_1(p), \tag{C7}$$

$$k_0(p) = -(1-p)\frac{\partial K_0(p)}{\partial p} + K_0(p). \tag{C8}$$

The polynomials in $\hat{p}$ from step 2 for treated and untreated observations form the estimates of $K_1(p)$ and $K_0(p)$, respectively. As both are polynomials, the derivatives of these expressions with respect to p are simple to compute.

(7) Using estimates of β_0 and β_1 from step 4, and $k_0(p)$ and $k_1(p)$ from step 6, calculate $\widehat{\text{MTE}}$ as follows for each value of $(x, \hat{p})$:

$$\widehat{\text{MTE}}(X = x, U_D = p) = x\hat{\beta}_1 - x\hat{\beta}_0 + \widehat{k_1(p)} - \widehat{k_0(p)}. \tag{C9}$$

(8) If desired, estimate standard errors for $\widehat{\text{MTE}}$ using a bootstrap procedure. In our estimation, we use 100 replications.

References

Borchert, J. R. (1987), *America's Northern Heartland*. University of Minnesota Press. [3]

Brinch, C. N., M. Mogstad, and M. Wiswall (2017), "Beyond LATE with a discrete instrument." *Journal of Political Economy*, 125, 985–1039. [4]

Dustmann, C. and J-S. Görlach (2016), "The economics of temporary migrations." *Journal of Economic Literature*, 54, 98–136. [1]

Dustmann, C. and A. Okatenko (2016), "Out-migration, wealth constraints, and the quality of local amenities." *Journal of Development Economics*, 110, 52–63. [1]

Fan, J. and I. Gijbels (1996), *Local Polynomial Modelling and Its Applications*. Chapman and Hall. [5]

Imbens, G. W. and J. D. Angrist (1994), "Identification and estimation of local average treatment effects." *Econometrica*, 62, 467–475. [4]

Ruggles, S., J. T. Alexander, K. Genadek, R. Goeken, M. B. Schroeder, and M. Sobek (2010), *Integrated Public Use Microdata Series: Version 5.0 [Machine-Readable Database]*. University of Minnesota, Minneapolis. [3]

Co-editor Christopher Taber handled this manuscript.

Manuscript received 4 September, 2017; final version accepted 13 September, 2018; available online 16 November, 2018.