Semiparametric Analysis of German East–West Migration – Facts and Theory

Michael C. BURDA[1]

Wolfgang HÄRDLE[2]

Marlene MÜLLER[2]

Axel WERWATZ[2]

[1] Institut für Wirtschaftstheorie II

[2] Institut für Statistik und Ökonometrie

[1+2] Sonderforschungsbereich 373

Wirtschaftswissenschaftliche Fakultät

Humboldt–Universität zu Berlin, Germany

1

Introduction

- German East-West migration in 1991
- microdata from the German Socio Economic Panel (GSOEP), $n = 3367$
- Generalized Linear Model (GLM) does not fit the data
- semiparametric Generalized Partial Linear Model (GPLM) reveals nonlinear influence of household income on migration propensity
- this nonlinear influence is compatible with the option value approach of Burda (95)

Classical Economic Theory

- income is key determinant of migration
- difference between income the host region W^W and income at home W^E at time t (1991: $t = 0$): $\Omega_t = W_t^W - W_t^E$
- forward-looking agent will consider expected net present value (ENPV) =

    EPV of income from migrating
  − EPV of income from not-migrating
  − fixed costs of migrating

- under standard assumptions ENPV is a linear function of current (=1991) income differential Ω_0.

Example:

Ω_t follows Brownian motion with drift ν :

$$d\,\Omega_t \;=\; \nu dt + \sigma\, dz_t$$

where $dz_t = \epsilon_t \sqrt{dt}$, $\epsilon_t \sim N(0,1)$.

$$\Rightarrow \text{ENPV} \;=\; V^m \;=\; \frac{1}{\delta}\,(\Omega_0 + \nu/\delta) \;-\; F$$

where δ is the rate of discount.

Marshallian decision rule

$$\begin{aligned} Y &= 1 \qquad \text{if } V^m > 0 \\ Y &= 0 \qquad \text{otherwise} \end{aligned}$$

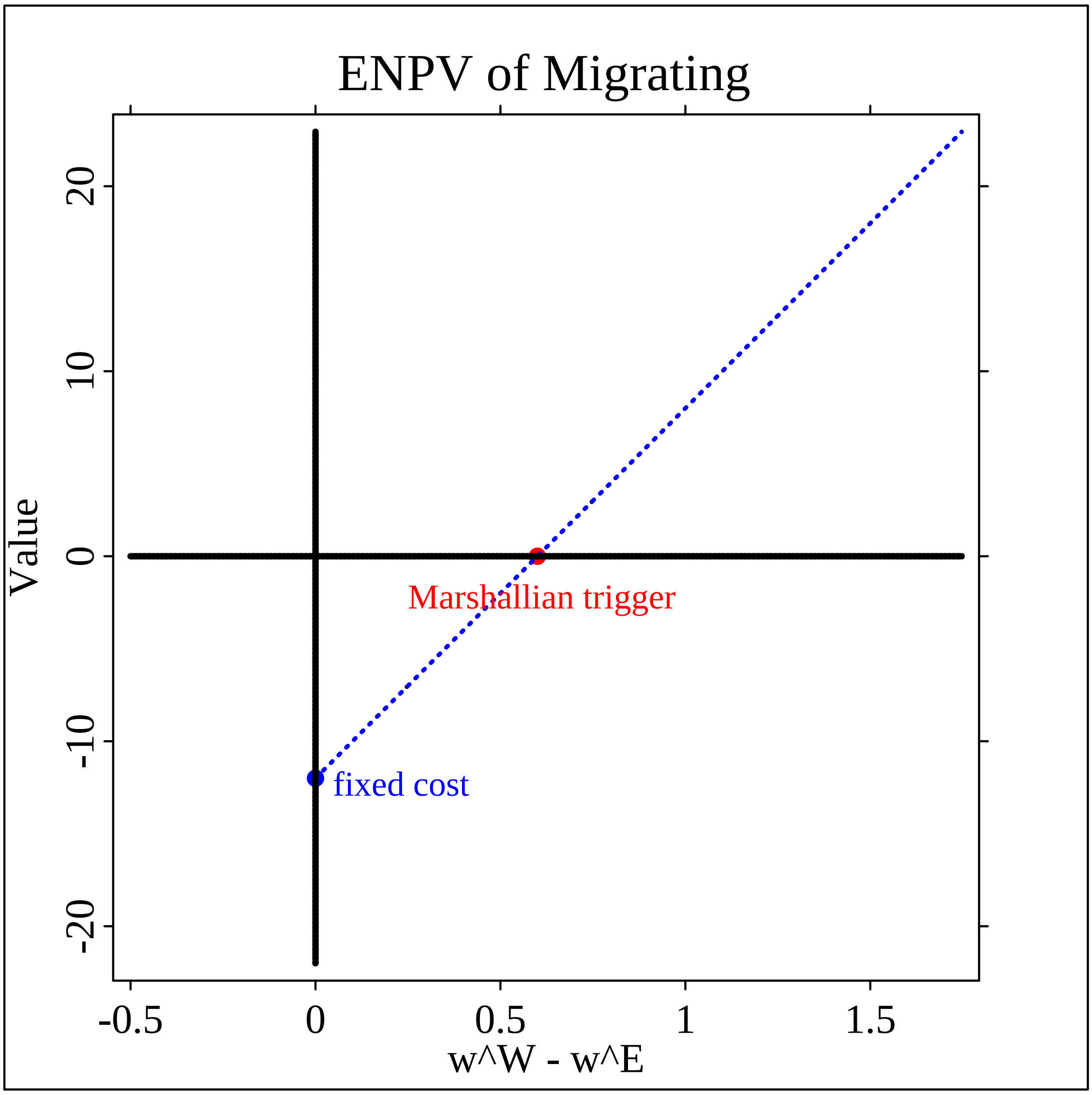


Figure 1: Marshallian theory of migration

The Data

- 3367 observations from GSOEP's 2nd East-German wave (spring of 1991)
- dependent variable Y: migration *propensity*
- measuring current income differential Ω_0: imputation is prone to error (self-selection, unemployed, out of the labor force) include income in East (W_0^E) only
- 11 explanatory variables
- All calculations were done in XploRe©. See `http://www.xplore-stat.de`

Summary statistics

		Mean	Expected Effect
Y	**migration intention**	.39	
X_1	**female**	.51	
X_2	**partner**	.85	$-$
X_3	**owner**	.32	$-$
X_4	**family/friends in west**	.85	$-$
X_5	**unemp./jobloss certain**	.20	$+$
X_6	**env. satisfaction**	3.9	$-$
X_7	**city size $<$ 10,000**	.52	
X_8	**city size 10-100,000**	.34	
X_9	**university degree**	.08	
X_{10}	**age**	39.4	$-$
	min: 18, max: 65		
X_{11}	**household income**	2189.5	
	min: 200, max: 4000		

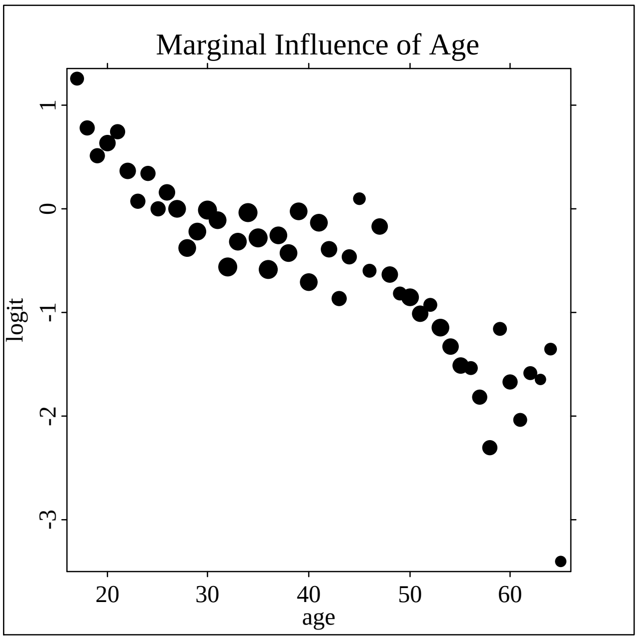
Parametric Estimation Results

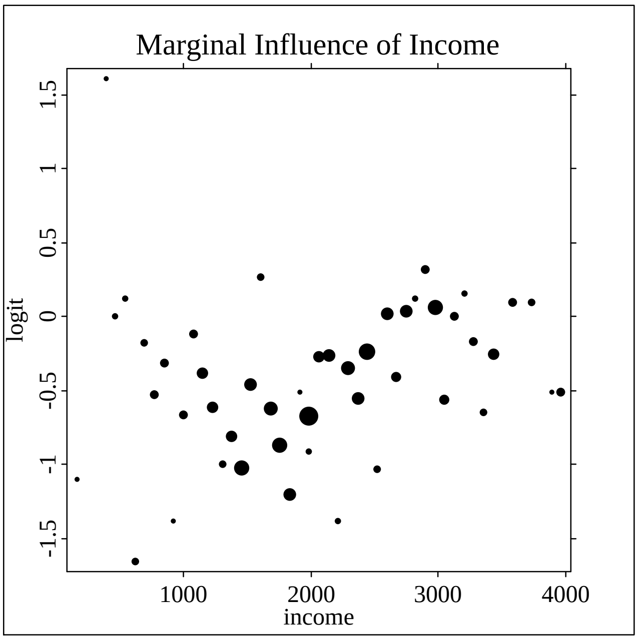
GLM (Logit) estimates of β in $E[Y|x] = 1/\{1 + \exp(-\beta^T x)\}$.

dependent variable: migration intention		
Variable	$\widehat{\beta}$	t
constant	1.864	7.74
female	-.233	-3.03
partner	-.325	-2.87
owner	-.576	-5.79
family/friends in west	.647	5.61
unemployed	.217	2.24
env. satisfaction	-.057	-3.52
city size $<$ 10,000	-.718	-5.69
city size 10-100,000	-.347	-2.91
university degree	.481	3.56
age	-.050	-14.89
household income	.0001202	2.22
sample size: 3367, log likelihood: -1992.7		

income & age: linear or nonlinear ?

age and income vs. the logits $\log\{\widehat{p}/(1-\widehat{p})\}$





Semiparametric Model

Latent-variable assumption

$$\text{GLM:} \quad Y = 1 \quad \text{if } Y^* = x^T\beta + \alpha\, t + \alpha_0 - u \quad > 0$$

$$\text{GPLM:} \quad Y = 1 \quad \text{if } Y^* = x^T\beta + m(t) - u \quad > 0$$

t: income in the East (W_0^E)

Distributional assumption

$$\text{GLM \& GPLM:} \quad F_{u|x,t}(\bullet) = \frac{1}{1 + \exp(-\bullet)}$$

GPLM:

$$E(Y|x,t) = \frac{1}{1 + \exp[-\{x^T\beta + m(t)\}]}$$

Semiparametric Estimation

- $\widehat{\beta}$ can be found for known m,
- $\widehat{m}$ can be found for known β.

Iterative algorithm (Link function !) employs:

- "usual" likelihood for β

$$\mathcal{L}(\beta) = \sum_{i=1}^{n} L\left\{x_i^T\beta + m_\beta(t_i); y_i\right\}$$

- "smoothed" likelihood for $m(t)$

$$\mathcal{L}^S\{m_\beta(t)\} = \sum_{i=1}^{n} K_h(t - t_i)\, L\left\{x_i^T\beta + m_\beta(t); y_i\right\}$$

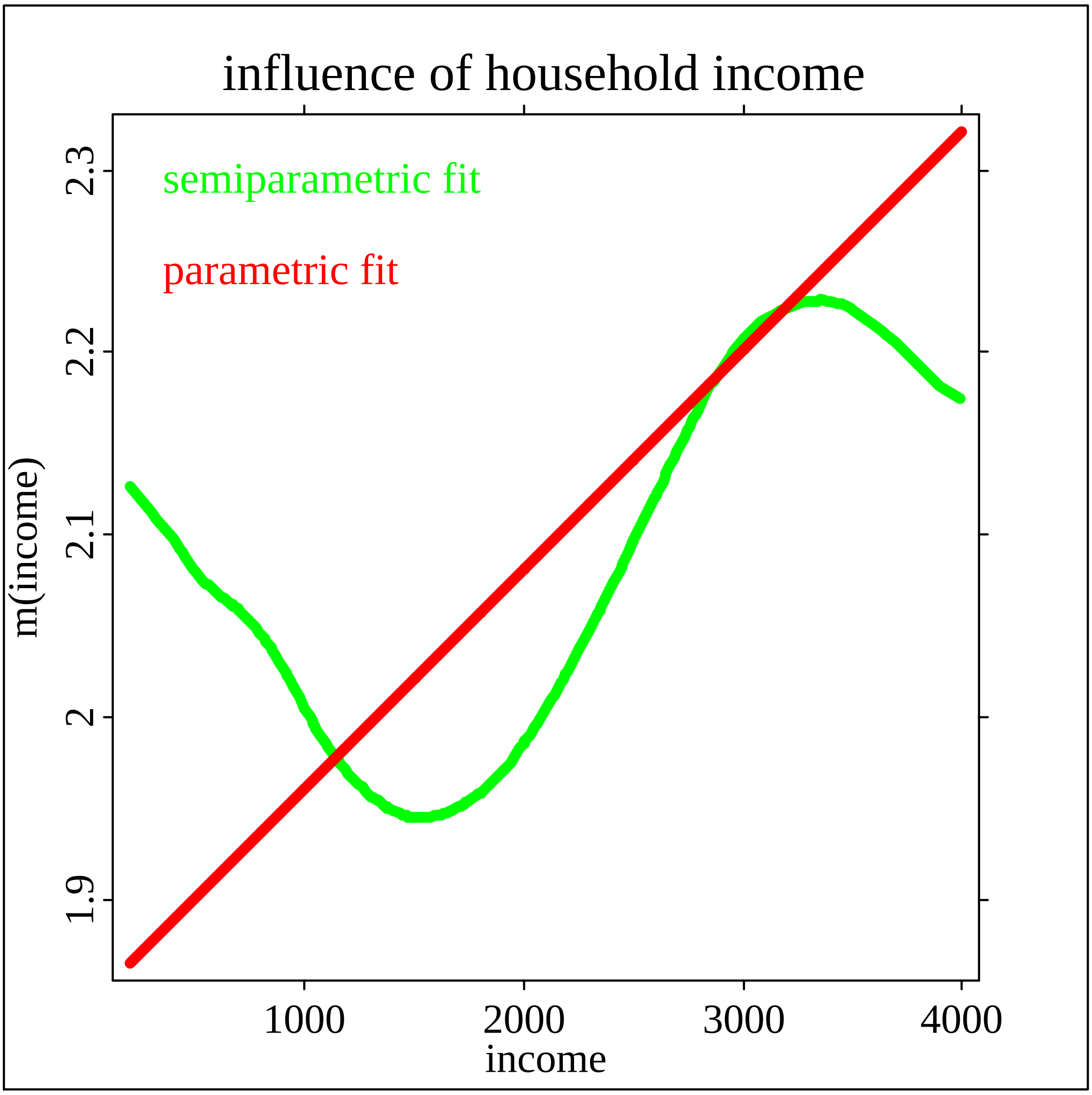
Severini & Staniswalis (1994), Severini & Wong (1992), Hastie & Tibshirani (1990), Speckman (1988)

Semiparametric Estimation Results

dependent variable: migration intention				
	GPLM		Logit	
Variable	$\hat{\beta}$	t	$\hat{\beta}$	t
female	-.238	-3.1	-.233	-3.0
partner	-.282	-2.4	-.325	-2.9
owner	-.569	-5.7	-.576	-5.8
family/friends in west	.640	5.5	.647	5.6
unemployed	.216	2.2	.217	2.2
env. satisfaction	.056	-3.5	-.057	-3.5
city size $<$ 10,000	-.689	-5.4	-.718	-5.7
city size 10-100,000	-.323	-2.7	-.347	-2.9
university degree	.471	3.5	.481	3.6
age	-.050	-14.9	-.050	-14.9
sample size: 3367, log likelihood: -1989.8, $h = 0.3$				

GPLM estimates are close to Logit counterparts

estimated influence of income: $\widehat{m}(t)$



Semiparametric Specification Testing

test that $m(t)$ is a linear function:

$$
\begin{aligned}
H_0 &: \quad m(t) = \alpha\, t + \alpha_0 \,, \\
H_1 &: \quad m(t) \text{ is an arbitrary smooth function,}
\end{aligned}
$$

Likelihood ratio test (Hastie & Tibshirani, 1990)

$$R = 2 \sum_{i=1}^{n} \{\, L(\widehat{\mu}_i, y_i) - L(\widetilde{\mu}_i, y_i)\,\}$$

semiparametric: $\widehat{\mu}_i = G\{x_i^T \widehat{\beta} + \widehat{m}(t_i)\}$

parametric: $\widetilde{\mu}_i = G\{x_i^T \widetilde{\beta} + \widetilde{\alpha}\, t + \widetilde{\alpha}_0\}$

! $\widehat{m}$ has a non–negligible smoothing bias

Modified likelihood ratio test

bias–corrected parametric estimate

$$\overline{m}(t_j)$$

from

$$\{G(x_i^T \widetilde{\beta} + \widetilde{\alpha}\, t_i + \widetilde{\alpha}_0), x_i, t_i\}, \quad i = 1, \ldots, n$$

modified LR statistic

$$R^M = 2 \sum_{i=1}^{n} \{ L(\widehat{\mu}_i, \widehat{\mu}_i) - L(\overline{\mu}_i, \widehat{\mu}_i) \}$$

where $\overline{\mu}_i = G\{x_i^T \widetilde{\beta} + \overline{m}(t_i)\}$

Härdle, Mammen & Müller (1996)

asymptotically equivalent

$$\widetilde{R}^M = \sum_{i=1}^{n} w_i \left\{ x_i^T (\widehat{\beta} - \widetilde{\beta}) + \widehat{m}(t_i) - \overline{m}(t_i) \right\}^2$$

with

$$w_i = \frac{[G'\{x_i^T \widehat{\beta} + \widehat{m}(t_i)\}]^2}{V[G\{x_i^T \widehat{\beta} + \widehat{m}(t_i)\}]}.$$

Asymptotic Normality

Under linearity hypothesis

(i) $R^M = \widetilde{R}^M + o_p(v_n)$,

(ii) $v_n^{-1}(R^M - e_n) \xrightarrow{D} N(0,1)$,

where

$$\begin{aligned} e_n &= \left\{ \lambda_T \cdot \int K(u)^2 \, du \right\} \{h_1 \ldots h_q\}^{-1}, \\ v_n^2 &= 2 \left[\lambda_T \int \{K \star K(u)\}^2 \, du \right] \{h_1 \ldots h_q\}^{-1}, \end{aligned}$$

Bootstrap works

It holds

$$d_K(R^{M*}, R^M) \xrightarrow{P} 0$$

where d_K denotes the Kolmogorov distance.

1. Generate samples $y_1^*, \ldots, y_n^*$ with

$$E^*(y_i^*) \quad = \quad G(x_i^T \widetilde{\beta} + \widetilde{\alpha}\, t_i + \alpha_0)$$

2. Calculate estimates based on the bootstrap samples and finally the test statistics R^{M*}. The quantiles of the distribution of R^M are estimated by the quantiles of the conditional distributions of R^{M*}.

Test Results

h	0.1	0.2	0.25	0.3	0.4
R	0.028	0.021	0.019	0.017	0.016
R^M	0.053	0.069	0.130	0.269	0.602
R^{M*}	0.015	0.005	0.005	0.005	0.010

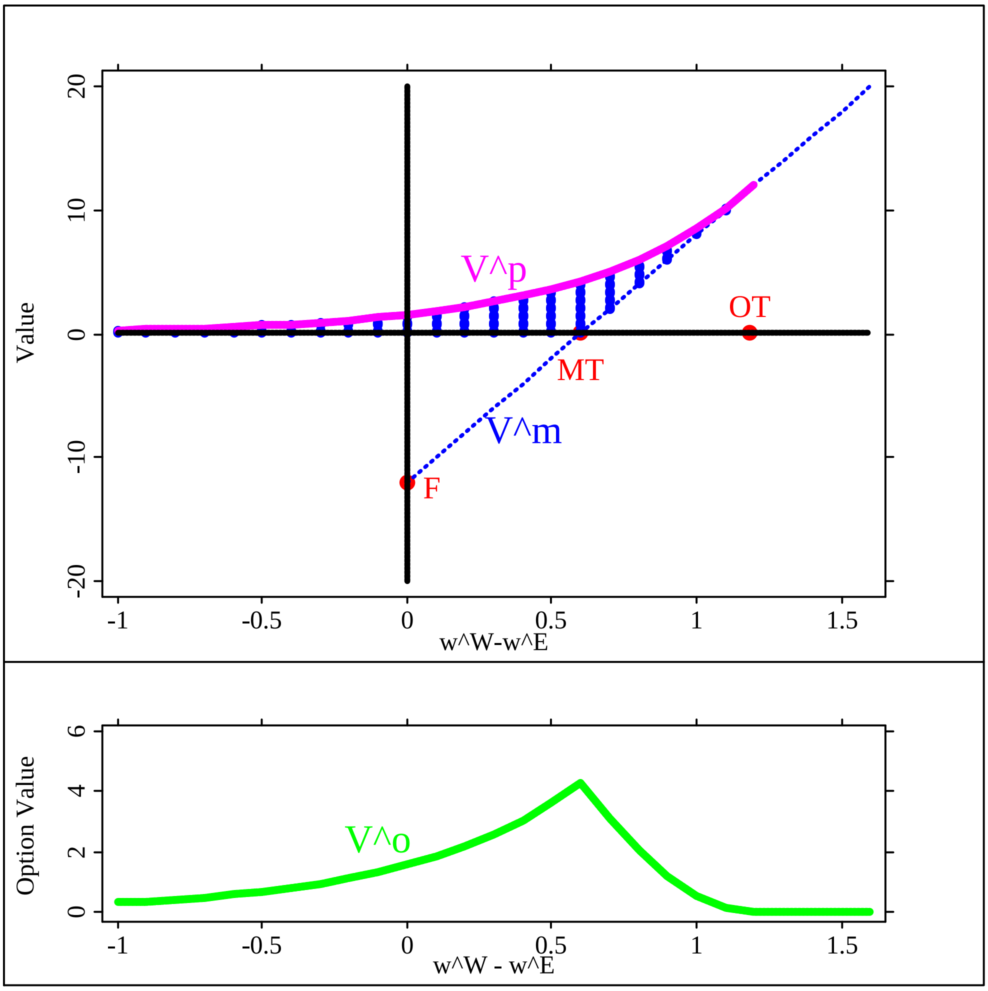
- clear rejection of the linearity hypothesis across all bandwidths for R and the bootstrapped R^{M*}.

- The normal approximation for R^M works bad for higher bandwidth levels (Müller, 1997)

Theoretical Explanation I

Option Value of the Migration Investment

- Marshallian theory: migration occurs now or never.
- Dixit and Pindyck (1994): postponement of the decision without forsaking it can be a valuable option
- delaying migration: more information can be acquired while fixed cost can be avoided
- migrating today means forgoing the opportunity to postpone migration option value of waiting V^o.
- $V^o =$ what one is willing to pay for the option to postpone the migration decision rather than having to decide now or never



V^p : expected net present value from postponing migration

V^m : expected net present value from migrating today

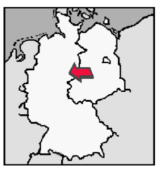
V^o : option value of waiting

Marshallian decision rule

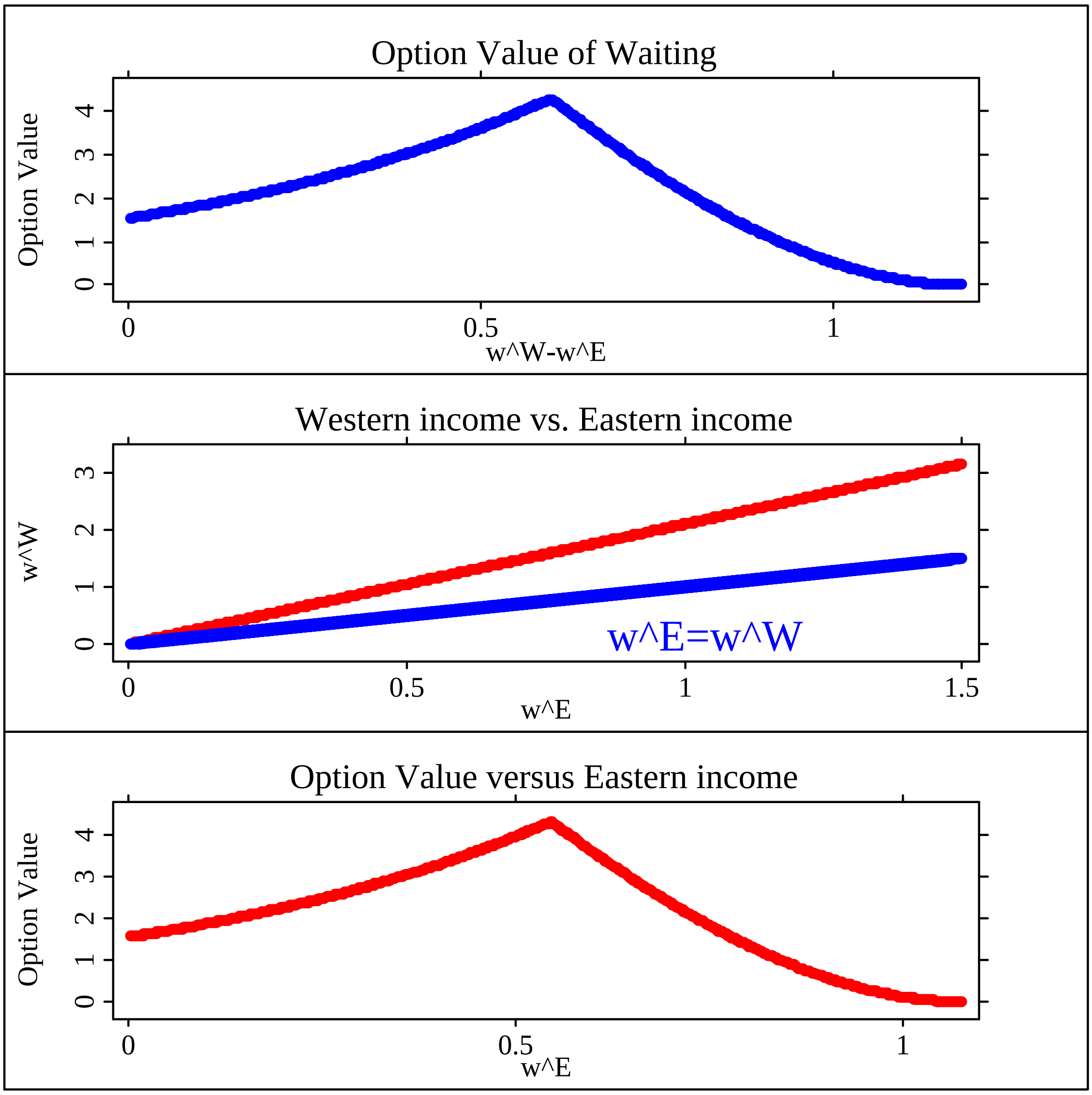
$$
\begin{aligned}
Y &= 1 \qquad \text{if } \frac{1}{\delta}(\Omega_0 + \nu/\delta) - F > 0 \\
Y &= 0 \qquad \text{otherwise}
\end{aligned}
$$

Option value decision rule

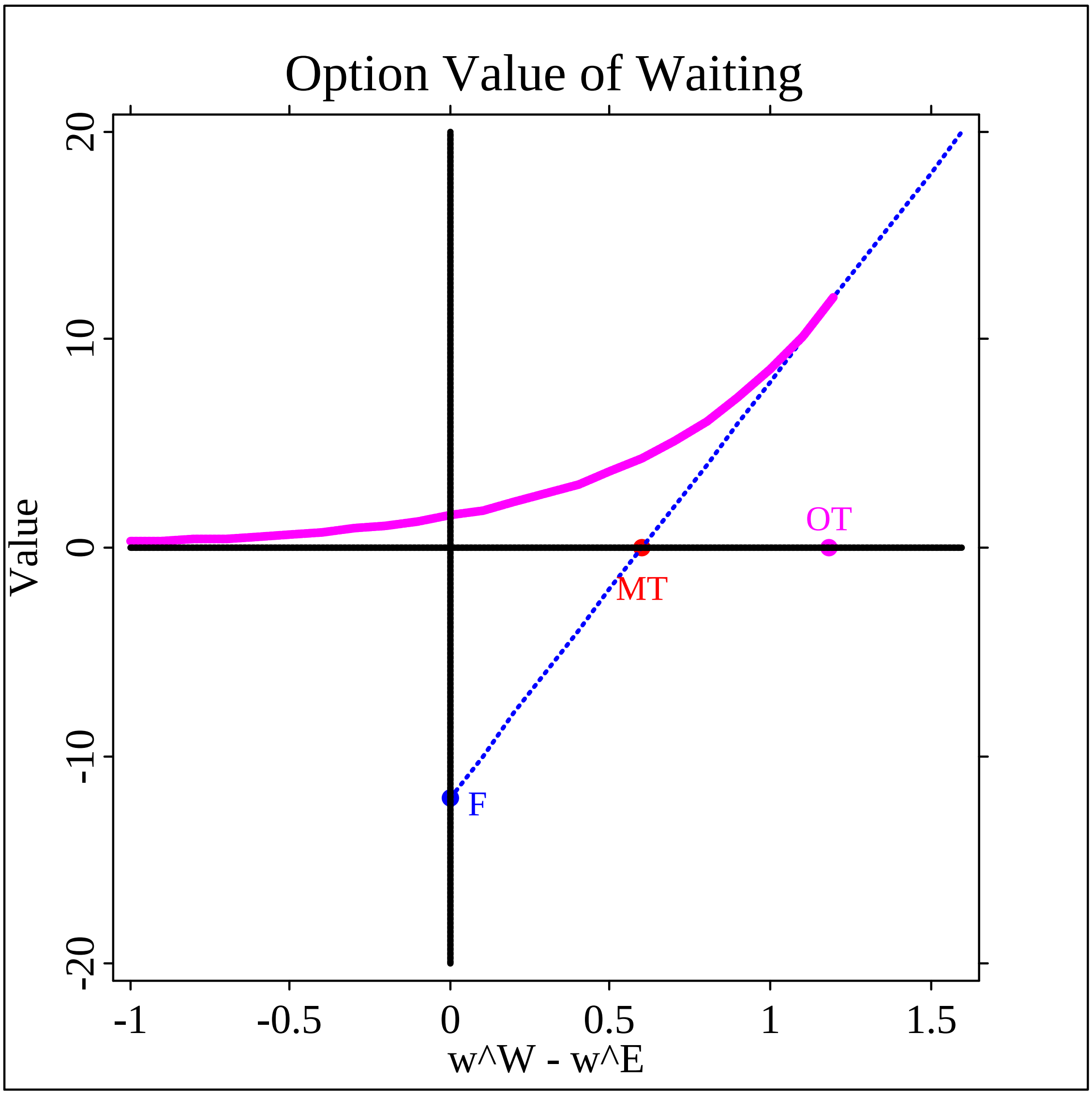
$$
\begin{aligned}
Y &= 1 \qquad \text{if } \frac{1}{\delta}(\Omega_0 + \nu/\delta) - F - V^o(\Omega_0) > 0 \\
Y &= 0 \qquad \text{otherwise}
\end{aligned}
$$



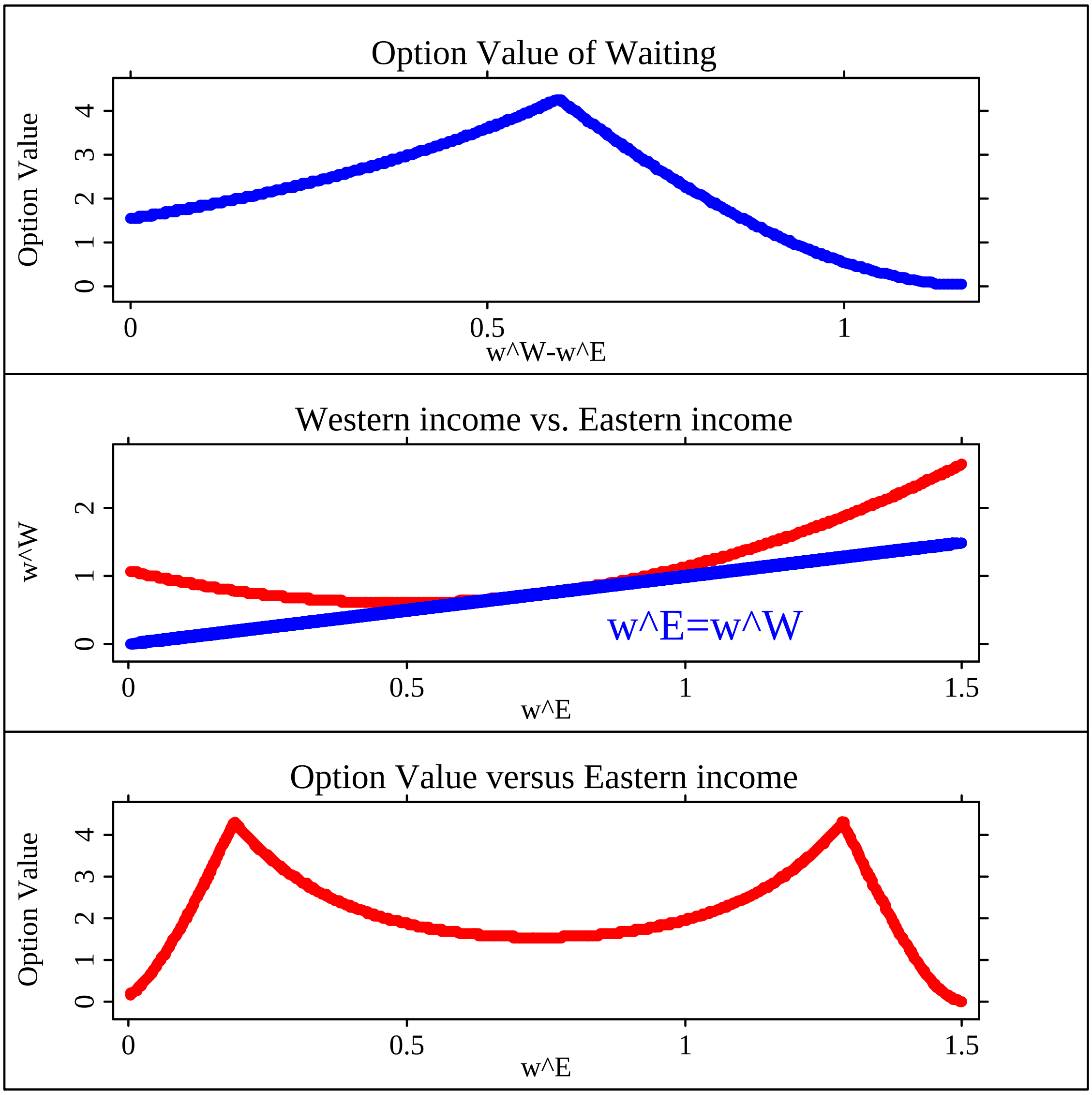
income *differential* versus income in East.



Theoretical Explanation II



income *differential* versus income in East.



Conclusions

- empirical analysis of the propensity to migrate using microdata from the GSOEP
- parametric GLM did not fit the data
- semiparametric GPLM fit produced U-shaped relation between income and migration propensity
- U-shaped relation significantly deviates from linearity
- estimated influence may be explained by a number of alternative determinants of migration, including the recently proposed option-value-of-waiting theory