How to use economic theory to improve estimators, with an application to labor demand and wage inequality

Maximilian Kasy Pirmin Fessler

February 29, 2016

Points of departure

1. Use of theory in empirical research
    - standard approach: positivist, imported from physics
    - testing / imposing theories
    - alternative: procedures that perform well
      if theories are approximately true
2. Larger agenda: econometric methods to better understand
    - income inequality, in particular wage inequality
    - changes in inequality
    - historical causes, policy counterfactuals, predictions
    - winners and losers; political economy

Points of departures continued

3. Empirical Bayes procedures in applied econometrics
    - achieve efficiency gains as in James-Stein shrinkage
    - avoid requirement for applied researchers to specify tuning parameters / priors
    - proposal: shrink towards theory, rather than arbitrary point 0
4. Impact of labor supply on wage inequality
    - many types of workers $\Rightarrow$ many regressors, few observations
    - conventional solution: parametric structural model
    - non-robust conclusions
    - proposed solution: shrinking towards structural model in data-dependent, optimal way

Literature

1. **Empirical Bayes**:
   Robbins (1956), James and Stein (1961), Efron and Morris (1973). Morris (1983), Laird and Louis (1987), Carlin and Gelfand (1990), Efron (2010).
2. **Labor - determinants of wage distribution**:
   Borjas et al. (1996), Autor et al. (1998), Autor et al. (2008), Card (2001), Card (2009), Boustan (2009), Autor and Dorn (2013)

Outline

1. Theory in empirical research
2. Proposed procedure to construct empirical Bayes estimator based on economic theory
3. Properties: Consistency, geometry
   Key result: characterization of risk function
4. Application to labor demand
   European wage inequality, EU-SILC data 2004-2013
5. Monte Carlo simulations
6. Conclusion and outlook

Warm-up – review of CES-production functions

Notation:

- types of workers $j = 1, \dots, J$,
  cross-section of labor markets $i = 1, \dots, n$
- wages w, labor supply N
- $Y_{ij} = \log(w_{ij})$, $X_{ij} = \log(N_{ij})$

Assumptions:

1. marginal productivity theory of wages:

$$w_{ij} = \frac{\partial f_i(N_{i1}, \dots, N_{iJ})}{\partial N_{ij}}$$

2. CES production function:

$$f_i(N_{i1}, \dots, N_{iJ}) = \left(\sum_{j=1}^{J} \gamma_j N_{ij}^{\rho} \right)^{1/\rho}$$

Wage equation

- This yields

$$w_{ij} = \frac{\partial f_i(N_{i1}, \dots, N_{iJ})}{\partial N_{ij}} = \left(\sum_{j'=1}^{J} \gamma_j N_{ij'}^{\rho} \right)^{1/\rho - 1} \cdot \gamma_j \cdot N_j^{\rho - 1}.$$

- relative wage between groups j and j' is equal to

$$\frac{w_{ij}}{w_{ij'}} = \frac{\gamma_j}{\gamma_{j'}} \cdot \left(\frac{N_{ij}}{N_{ij'}} \right)^{\rho - 1}.$$

- Taking logs yields

$$Y_{j,i} - Y_{j',i} = \log(\gamma_j) - \log(\gamma_{j'}) + \beta_0 \cdot (X_{j,i} - X_{j',i}),$$

where $\beta_0 = \rho - 1$.

Example: Impact of migration on wage inequality

- Literature: Estimate CES-production function model, consider historical counterfactual of no immigration.
- "Migration increased inequality"
  - Borjas et al. (1996)
  - CES-model with 4 types, by education
  - national economy, time series variation
- "Migration did not increase inequality"
  - Card (2001), Card (2009)
  - nested CES, 2 education types, natives vs migrants (justified by pre-tests)
  - cross-city, Bartik-type instrument

⇒ Conclusions depend on functional form choices!

Theory in empirical economics

- Structural models seem to cause non-robustness, possibly inconsistency.
  $\Rightarrow$ Should we use theory in empirical research at all?

The positivist ideal:

- Follow the example of physics.
- Develop theories which
  1. are assumed to be universally true, and
  2. have testable implications.
- Maintain these theories while they have not been rejected by statistical tests.
- When they have been rejected, replace them with new theories that are consistent with all available evidence.

The reality of economic theory

- We have no theories that are even approximately universally true.
- People don't universally – or even consistently in well defined contexts –
  - maximize utility,
  - discount exponentially,
  - maximize expected utility under risk,
  - play Nash equilibrium,
  - act as price takers on markets, ...
- Even less
  1. do people maximize utility with additive EV1 errors,
  2. does aggregate production follow a CES production function with 3 inputs, ...
- All of these theories can be, and have been, rejected.

What to do?

Several options:

1. Ignore this, keep following the positivist ideal, argue that theories don't actually have to be true.
   (Wasn't there something about Billiard players?)
2. Forget about economic theory, just try to do good statistics / mostly harmless econometrics.
3. Try to find a middle ground that makes reasonable use of theory.

An attempt at a middle ground

- Shrink “towards theory”
- Advantages:
  - Improves estimator performance if theory is (approximately) true.
  - Is not dogmatic – yields consistent estimates either way.
- Bayesian interpretation: (improper) priors that put low weight on parameter values deviating a lot from theory.
- Can do empirical Bayes version – avoids critique of subjectivism / arbitrary choice of tuning parameters.
- Might yield James-Stein type shrinkage informed by theory.
- Coming next: an implementation of this program in the context of labor demand.

The parametric empirical Bayes approach

- Parameters η, hyper-parameters θ
- **Model**:
$$Y|\eta \sim f(Y|\eta)$$
- **Family of priors**:
$$\eta \sim \pi(\eta|\theta)$$
- **Marginal density** of Y:
$$Y|\theta \sim g(Y|\theta) := \int f(Y|\eta)\pi(\eta|\theta)d\eta$$
- Estimation of hyperparameters: marginal MLE
$$\widehat{\theta} = \underset{\theta}{\operatorname{argmax}}\ g(Y|\theta).$$
- Estimation of η:
$$\widehat{\eta} = E[\eta|Y, \theta = \widehat{\theta}]$$

Simplified setup for estimation with theory

1. Preliminary, unrestricted estimator:

$$\widehat{\beta} \sim N(\beta, V)$$

2. Restriction implied by theory:

$$\beta = \beta_0 \cdot M$$

Conventional approaches:

- structural estimation
- unrestricted estimation
- pre-testing

An empirical Bayes approach for our setup

- **Model**:
  unrestricted estimator as sufficient statistic;
  asymptotic approximation:

$$\widehat{\beta} \sim N(\beta, V)$$
$$\widehat{V} \cdot V^{-1} \to^p I.$$

- **Family of priors**:
  coefficients = structural model + noise of unknown variance

$$\beta = \beta_0 \cdot M + \zeta$$
$$\zeta_j \sim^{iid} N(0, \tau^2),$$

- Parameters η, hyper-parameters θ:

$$\begin{aligned}
\eta &= (\beta, V) \\
\theta &= (\beta_0, \tau^2, V) \\
\widehat{\beta}|\eta &\sim N(\beta, V) \\
\beta|\theta &\sim N(\beta_0 \cdot M, \tau^2 \cdot I).
\end{aligned}$$

- **Marginal density** of Y:

$$\widehat{\beta}|\theta \sim N(\beta_0 \cdot M, \Sigma(\tau^2, V))$$

where

$$\Sigma(\tau^2, V) = \text{Var}\left(\widehat{\beta}\middle|\theta\right) = \tau^2 \cdot I + V.$$

Solving for the estimator

- Hyperparameters: MLE for the marginal likelihood,

$$(\widehat{\beta}_0, \widehat{\tau}^2) = \underset{b_0, t^2}{\operatorname{argmin}} \log\left(\det(\Sigma(t^2, \widehat{V}))\right) + (\widehat{\beta} - b_0 \cdot M)' \cdot \Sigma(t^2, \widehat{V})^{-1} \cdot (\widehat{\beta} - b_0 \cdot M).$$

- Parameter of interest β:

$$\widehat{\beta}^{EB} = \widehat{\beta}_0 \cdot M + \left(I + \frac{1}{\widehat{\tau}^2}\widehat{V}\right)^{-1} \cdot (\widehat{\beta} - \widehat{\beta}_0 \cdot M).$$

Empirical Bayes confidence sets

- Require
$$P(\beta \in C|\theta) \geq 1-\alpha$$
- hybrid of
  1. frequentist coverage
$$P(\beta \in C|\eta) \geq 1-\alpha$$
  2. Bayesian pre-posterior analysis
$$P(\beta \in C) \geq 1-\alpha$$
- cf. Laird and Louis (1987).

Implementation of confidence sets

1. Generate $r = 1, \dots, R$ i.i.d. draws $\widehat{\beta}_r$ from $N\left(\widehat{\beta}, \widehat{V}\right)$.
2. For each r, obtain $\widehat{\theta}_r = (\widehat{\beta}_{0,r}, \widehat{\tau}_r^2)$, $\widehat{\beta}_r^{EB}$
3. Let
$$V_r^{EB} = \text{Var}(\beta | \widehat{\beta} = \widehat{\beta}_r, \theta = \widehat{\theta}_r) = \left(I + \tfrac{1}{\widehat{\tau}^2}\widehat{V}\right)^{-1} \cdot \widehat{V}.$$
4. Consider the mixture distribution
$$M\left(\beta | \widehat{\beta}, \widehat{V}\right) := \frac{1}{R}\sum_r N\left(\widehat{\beta}_r^{EB}, V_r^{EB}\right).$$
5. Confidence intervals: bounded by quantiles of $M\left(\beta | \widehat{\beta}, \widehat{V}\right)$.

Aside: other uses of this framework in labor economics

- Many interesting recent papers estimate large number of fixed effects:
  - Location effects and intergenerational mobility: Chetty and Hendren (2015)
  - Teacher Effects: Chetty et al. (2014)
  - Worker and firm effects: Card et al. (2012)
  - Judge effects: Abrams et al. (2012)
- Our framework immediately applies to these settings:
  - Take $\widehat{\beta}$ as the OLS FE estimates.
  - Take $\widehat{V}$ as the appropriate (heteroskedasticity robust, clustered) variance matrix.
  - Assume $\beta_i \sim^{iid} N(\beta_0, \tau^2)$.
    or $\beta_i | W_i \sim N(W_i \cdot \gamma_0, \tau^2)$
  - All the following results apply.

Properties of our estimator

1. Consistent conditional on any value of β
2. Counterfactual predictions driven by data, whenever these are informative
3. Lower mean squared error (MSE) than unrestricted estimation for most of the parameter space.
   cf. James-Stein shrinkage!
4. Lower MSE than structural estimation for modest violations of theory.

Proposition (Consistency)

Suppose that $\widehat{\beta} \to^p \beta$ and that $\widehat{V} \to^p 0$. Then

$$\widehat{\beta}^{EB} \to^p \beta$$

as sample size n goes to infinity.

Data-driven predictions

- Direction where unrestricted estimates are precise:

$$x \cdot \widehat{V} \cdot x' \approx 0,$$

- Then

$$x \cdot \widehat{\beta}^{EB} = x \cdot \left[\widehat{\beta} + \widehat{V} \cdot \left(\widehat{\tau}^2 \cdot I + \widehat{V} \right)^{-1} \cdot (\widehat{\beta}_0 \cdot M - \widehat{\beta}) \right] \approx x \cdot \widehat{\beta}.$$

Geometry

- Empirical Bayes maps $\widehat{\beta} \to \widehat{\beta}^{EB}$
- Geometry of this mapping?
- Simplifying assumptions (for exposition):
  1. diagonal V (just a change of coordinates)
  2. $\beta_0 = 0$ (general case discussed in paper)
- Formally:

$$\widehat{\beta}|\beta \sim N(\beta, \text{diag}(v))$$
$$\beta|\tau^2 \sim N(0, \tau^2 \cdot I),$$

Empirical Bayes in this simplified setting

- $\widehat{\beta}^{EB}$ given $\widehat{\tau}^2$:

$$\widehat{\beta}^{EB} = \text{diag}\left(\frac{\widehat{\tau}^2}{\widehat{\tau}^2 + v_k}\right) \cdot \widehat{\beta}.$$

- Does not directly give mapping $\widehat{\beta} \to \widehat{\beta}^{EB}$, since $\widehat{\tau}^2$ depends on $\widehat{\beta}$
- FOC for $\widehat{\tau}^2$:

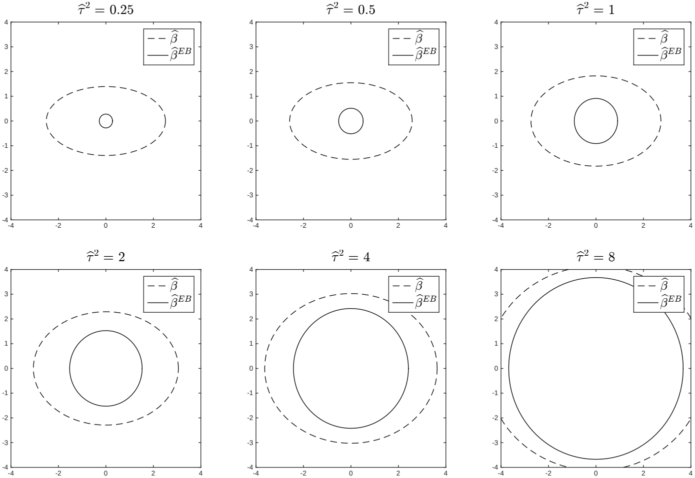
$$\sum_k \frac{1}{\widehat{\tau}^2 + v_k} = \sum_k \frac{\widehat{\beta}_k^2}{(\widehat{\tau}^2 + v_k)^2}.$$

Suppose $\widehat{\tau}^2$ is given

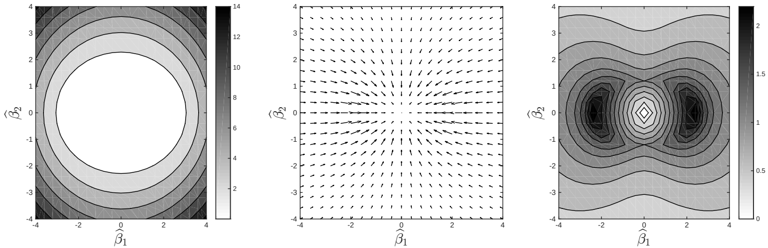
1. What's the set of $\widehat{\beta}$ yielding this $\widehat{\tau}^2$? **Ellipsoid** with semi-axes of length
$$(\widehat{\tau}^2 + v_k) \cdot \sqrt{\sum_{k'} \frac{1}{\widehat{\tau}^2 + v_{k'}}}.$$
2. What's the corresponding set of $\widehat{\beta}^{EB}$? **Circle** (!) of radius
$$\widehat{\tau}^2 \cdot \sqrt{\sum_{k'} \frac{1}{\widehat{\tau}^2 + v_{k'}}}.$$
3. $\widehat{\tau}^2$ is 0 inside ellipsoid with semi-axes of length
$$v_k \cdot \sqrt{\sum_{k'} \frac{1}{v_{k'}}}.$$

Fixing $\widehat{\tau}^2$

$\dim(\beta) = 2,\ v_1 = 2,\ v_2 = 1$



The mapping from $\widehat{\beta}$ to $\widehat{\tau}^2$ and $\widehat{\beta}^{EB}$



$\widehat{\tau}^2$, $\widehat{\beta}^{EB} - \widehat{\beta}$, and its length as a function of $\widehat{\beta}$

Two objective functions for τ^2

1. Expected marginal log likelihood of τ^2, given β:

$$ELLH(\tau^2) := \frac{1}{K} \cdot \sum_k \left(\log(\tau^2 + v_k) + \frac{\beta_k^2 + v_k}{\tau^2 + v_k} \right)$$

2. Mean squared error of (empirical) Bayes, given β:

$$MSE(\tau^2) := \sum_k \left[\left(\frac{\tau^2}{\tau^2 + v_k} \right)^2 \cdot v_k + \left(\frac{v_k}{\tau^2 + v_k} \right)^2 \cdot \beta_k^2 \right].$$

First order conditions

1. Maximizing expected marginal log likelihood of τ^2:

$$\sum_k \frac{1}{(\tau^2 + v_k)^2} \left(\tau^2 - \beta_k^2\right) = 0$$

2. Minimizing mean squared error of empirical Bayes:

$$\sum_k \frac{v_k^2}{(\tau^2 + v_k)^3} \cdot \left(\tau^2 - \beta_k^2\right) = 0$$

Theorem (Asymptotic risk)

Let

$$\widehat{\tau}^2 = \underset{\tau^2}{\operatorname{argmax}}\ LLH(\tau^2)$$

$$\overline{\tau}^2 = \underset{\tau^2}{\operatorname{argmax}}\ ELLH(\tau^2)$$

Suppose random effects setup: $(\widehat{\beta}_i, \beta_i, v_i)$ *are jointly i.i.d. with finite variance.*

Then, as $K \to \infty$*,*

$$SE(\widehat{\tau}^2) - MSE(\overline{\tau}^2) \to 0, \tag{1}$$

in probability and in L^1*.*

Characterizing the risk function, in generalization of James-Stein!

(Un)restricted estimation of labor demand

- k_j: aggregate type k corresponding to type j
- $\tilde{X}_{ik} = \log(\tilde{N}_{ik})$
- $X_{ij} = \log(N_{ij}/\tilde{N}_{ik_j})$
- Restricted model: 2-type CES labor demand, observing wages for
- Unrestricted model: allow wages to additionally depend on distribution across sub-types
-

$$
\begin{aligned}
Y_{ij} - Y_{i1} &= (\gamma_j - \gamma_1) && + \beta_0 \cdot (\tilde{X}_{ik_j} - \tilde{X}_{i1}) + (\varepsilon_{ij} - \varepsilon_{i1}), \\
Y_{ij} - Y_{i1} &= (\gamma_j - \gamma_1) + \sum_{j'} \delta_{jj'} X_{ij'} && + \beta_0 \cdot (\tilde{X}_{ik_j} - \tilde{X}_{i1}) + (\varepsilon_{ij} - \varepsilon_{i1}).
\end{aligned}
$$

Adapting our general setup

- Family of priors: $\beta = (\beta_{j,j'}) \sim^{iid} N(0, \tau^2)$
- Variance of vectorized $\delta = \Delta \cdot \beta$:

$$\text{Var}\left(\delta_{\uparrow}\right) = \tau^2 \cdot P \otimes I_J,$$

where $P = \Delta \cdot \Delta' = I_{J-1} + E$.

- Thus:

$$\Sigma(\tau^2, V) = \text{Var}\left((\widehat{\delta}_{\uparrow}, \widehat{\beta}_1)\right) = \begin{pmatrix} \tau^2 \cdot P \otimes I & 0 \\ 0 & 0 \end{pmatrix} + V.$$

- Estimating hyper-parameters:

$$(\widehat{\beta}_0, \widehat{\tau}^2) = \underset{b_0, t^2}{\text{argmin}} \log\left(\det(\Sigma(t^2, \widehat{V}))\right) + (\widehat{\delta}_{\uparrow}, \widehat{\beta}_1 - b_0)' \cdot \Sigma(t^2, \widehat{V})^{-1} \cdot (\widehat{\delta}_{\uparrow}, \widehat{\beta}_1 - b_0)'.$$

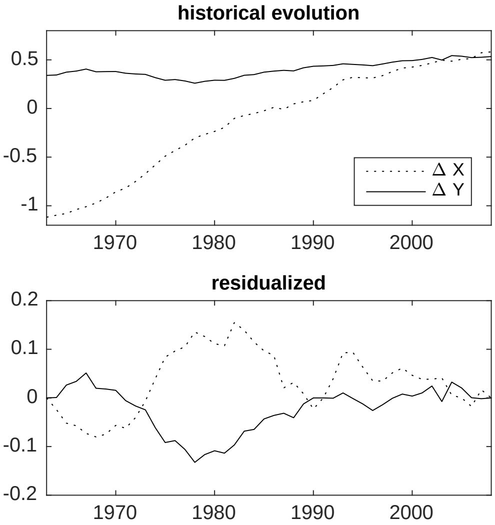
Empirical application 1: USA

- Much-studied datasets: American Community Survey (ACS), Current Population Survey (CPS).
- Following Acemoglu and Autor (2011), build national time series.
- Years 1963-2008 using the march CPS.
- Following Borjas et al. (2012), build state-level panel.
- Years 1960, 1970, 1980, 1990, and 2000 using the CPS, and 2006 using the ACS.

Sample and variable definitions

- Sample:
  - aged between 25 and 64,
  - less than 49 years of potential experience,
  - no self-employed or institutionalized workers.
- Labor supply: total hours worked per type.
- Average log wages: male full-time workers.
- Main analysis classifies workers into eight types:
  - by education (high school dropouts, high school graduates, some college, and college graduates),
  - and by potential experience (less than 20 years, and 20 years or more).

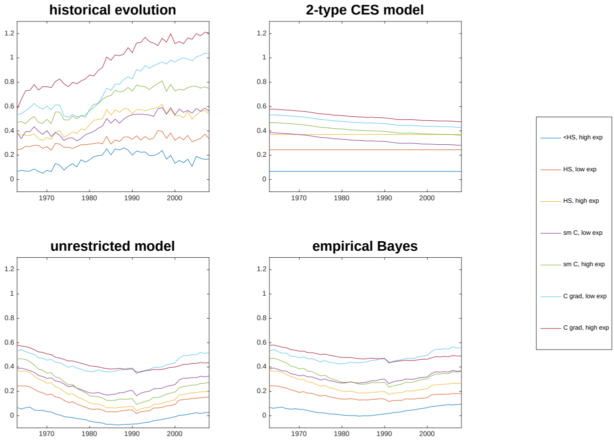
Log relative wages in the US – 2 types of workers



Summary of findings

- Estimate of 2-type inverse elasticity:
  1. Time-series with trend: -0.64 (replicating lit)
  2. State panel with FEs: -0.06, standard error 0.04.
- Model fit $\widehat{\tau}^2$ for 2-type model: not that great.
- Next slide:
  - counterfactual wage evolution,
  - based on national time series of labor supply,
  - with alternative estimates of labor demand using panel.

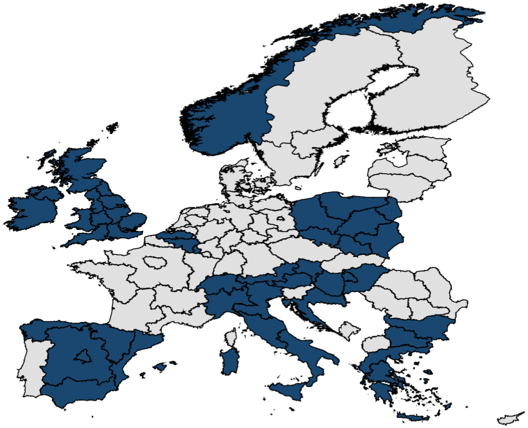
Log relative wages in the US – actual evolution and counterfactual changes



Empirical Application 2: Europe

- EU-SILC (2004-2013)
- NUTS1-level dataset
- Construct residual wages and labor supply for different labor types.
- Again following literature in terms of variable definitions.

Available regions



- Dark: used in our analysis
- Light: in EU-SILC, but not all variables available ⇒ excluded

Dependent variable: residual log wages Y_{ijt}

- Region i, type j, year t
- So far: $Y_{ijt} =$ log wage
- To control for changes of **composition** within types: $Y_{ijt} =$ residual of regression of log wage on observed demographics
- We control for age, age squared, age cubed, fulltime and lowtime dummies, labor types (education and potential experience), and interactions
- Estimated for every year separately, including all workers, with person level microdata

$\Rightarrow$ Generate mean residual wages for each type, males above age 24

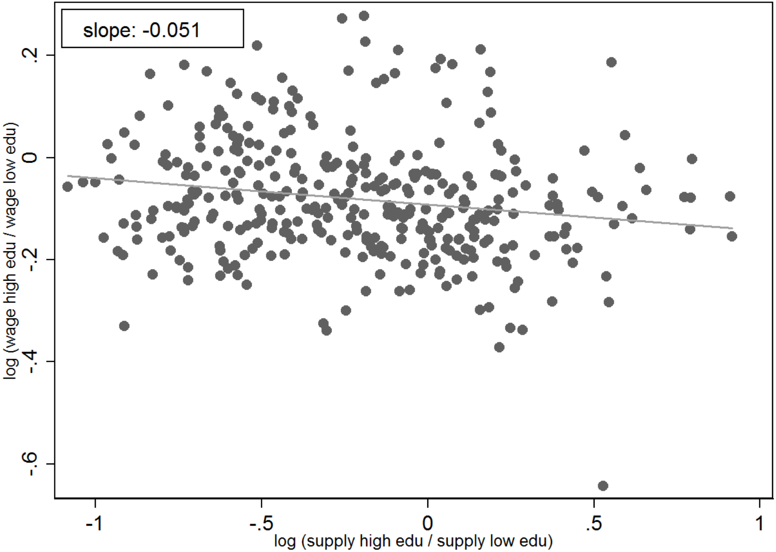
Regressors: log labor supply X_{ijt}

4 definitions of labor supply:

1. Efficiency weighted: sum of estimated log wages from cross-sectional wage regression (cf. previous slide)
2. Number of individuals
3. Number of hours worked
4. Efficiency weighted hours: estimated log wages from first stage $\times$ hours

Descriptive statistics, year 2010

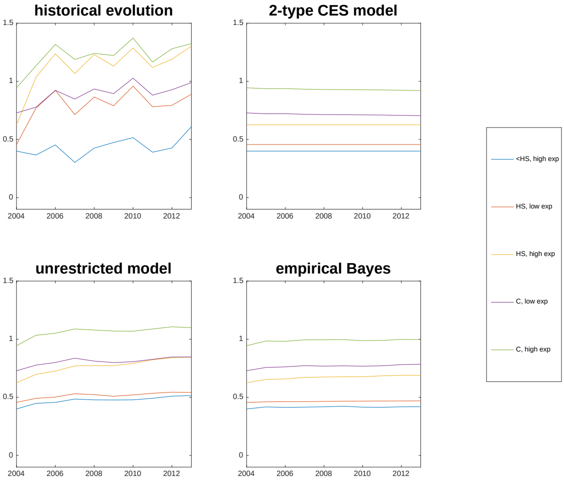
	All workers		Male over 24	
	Low edu	High edu	Low edu	High edu
Share female	0.44	0.54	0.00	0.00
	(0.002)	(0.003)	(0.000)	(0.000)
Mean age	41.5	40.6	43.2	41.8
	(0.052)	(0.065)	(0.065)	(0.097)
Mean workhours	37.8	38.6	41.3	41.7
	(0.044)	(0.061)	(0.050)	(0.084)
Share fulltime (>38h)	0.67	0.64	0.82	0.76
	(0.002)	(0.003)	(0.002)	(0.004)
Share migrant	0.07	0.07	0.07	0.07
	(0.001)	(0.002)	(0.002)	(0.002)
Number of Obs.	51,455	26,782	25,612	11,825


slope: -0.051
.2
0
-.2
-.4
-.6
log (wage high edu / wage low edu)
-1
-.5
0
.5
1
log (supply high edu / supply low edu)

Summary of findings

- Estimate of 2-type inverse elasticity: -0.05, standard error 0.07.
- Very similar to US panel!
- Model fit $\widehat{\tau}^2$ for 2-type model: again not that great.
- Next slide:
  - counterfactual wage evolution,
  - based on national time series of labor supply for Austria,
  - with alternative estimates of labor demand using panel.

Log relative wages in Austria – actual evolution and counterfactual changes



Some Monte Carlo evidence

- Comparing mean squared error of
    1. structural
    2. unrestricted
    3. empirical Bayes

  estimators.
- Two sets of simulations
    1. Conditional on θ:
       $\beta =$ structural model + random noise
    2. Conditional on η:
       β fixed

MSE relative to empirical Bayes conditional on θ

design parameters					MSE relative to empirical Bayes estimation		
n	J	σ^2	β_0	τ^2	structural	unrestricted	emp. Bayes
50	16	1.0	1.0	0.2	0.83	1.20	1.00
50	16	0.5	1.0	0.2	1.55	1.15	1.00
200	16	0.5	1.0	0.2	7.76	1.04	1.00
200	4	1.0	1.0	0.5	7.92	1.10	1.00
200	4	0.5	1.0	1.0	30.54	1.03	1.00

$$\beta = (\beta_{j,j'}) = \beta_0 \cdot M + \zeta$$
$$\zeta_{j,j'} \sim^{iid} N(0, \tau^2)$$

MSE relative to empirical Bayes conditional on η

design parameters						mean squared error		
n	J	σ^2	β_{00}	β_{01}	β_{02}	structural	unrestricted	emp. Bayes
200	4	1.0	1.0	1.0	1.0	0.18	1.47	1.00
200	4	1.0	1.0	1.0	6.0	15.04	1.03	1.00
200	4	1.0	0.0	1.0	6.0	19.37	1.01	1.00

$$\beta = \beta_{00} \cdot M_{J0} + \beta_{01} \cdot M_{J1} + \beta_{02} \cdot M_{J2},$$

- $M_{J0} = M$ in the first $J/4$ columns, zero elsewhere,
- $M_{J2} = M$ in the last $J/4$ columns, and zero elsewhere,
- $M_{J1} = M$ in the middle $J/2$ columns, and zero elsewhere.

Summary

- Object of interest: regressions with many regressors $\beta_{jj'}$, one for each pair of types of labor
- Restrictions of structural model:

$$\beta = \beta_0 \cdot M$$

- Structural model: inconsistent
  Unrestricted model: high variance
- Proposed solution: Empirical Bayes

$$\widehat{\beta}|\eta \sim N(\beta, V)$$
$$\beta|\theta \sim N(\beta_0 \cdot M, \tau^2 \cdot I).$$

Thanks for your time!