Deconvolution in Networks
Estimating Distributions of Firm and Worker Effects

Stéphane Bonhomme
University of Chicago

Adam Drozynski
University of Chicago

Thibaut Lamadon
University of Chicago

Preliminary Work!

DSE 2024 Conference

University of Wisconsin Madison, August 6 2024

Heterogeneous effects

- Linear regressions with fixed effects are increasingly used in economics.

- Leading examples are firm and worker effects in wage regressions, and neighborhood effects for economic mobility.

- These settings have a network structure: effects are identified by exploiting movements across firms or neighborhoods.

- The literature provides methods to recover the variance of effects (and covariances & regression coefficients).

- We study the estimation of general distributional features of the effects in such contexts, including the densities of effects.

Main example: workers and firms

• Abowd, Kramarz and Margolis (1999, AKM) study how worker and firm heterogeneity contribute to wage dispersion.

• The researchers are interested in the coefficients associated with worker and firm indicators.

• Questions: How dispersed are the firm effects ("firm premia")? How correlated are worker and firm effects ("sorting")?

• Recovering worker and firm effects requires exploiting movements between firms.

• Hence, the estimates of the effects depend on the network of employment relationships between workers and firms (and on how well connected it is).

Workers and firms (cont.)

• Economic models suggest that firm and worker effects are:

-correlated in complex ways, and

-correlated to the underlying network of employment relationships.

• To illustrate, we consider a Burdett-Mortensen job search model with firm heterogeneity.

• Our goal is to devise methods to recover distributions of worker and firm effects without restricting the conditional distribution of effects given the network.

Beyond variances & covariances

• Methods to estimate variances and covariances of effects are available (Andrews *et al.*, 2008, Kline *et al.*, 2020).

• However, many important economic quantities depend on distributions:

-An increase in the variance of firm effects has different implications depending on how the skewness changes, how tail probabilities evolve...

-Documenting the joint density of worker and firm effects sheds light on sorting along the distribution.

-Comparing firm effects before and after a job move has implications for search models.

• Today: we focus on marginal densities (but the method can be applied to bivariate densities too).

Another example: neighborhood effects

• Chetty and Hendren (2018) estimate the effects of neighborhoods in the US (such as counties or commuting zones) on income at adulthood.

• The times spent by the child in every neighborhood, which they refer to as neighborhood "exposures", are key covariates in the model.

• The exposure-time research design implies that the estimate of neighborhood j's effect depends on the outcome data on other neighborhoods j'.

• The authors rely on this design to estimate the causal effect of neighborhoods, under the assumption that the age at which children move across neighborhoods does not directly affect adult outcomes.

Related literature

- Nonparametric deconvolution (Stefanski and Carroll, 1990, Fan, 1991, Delaigle and Meister, 2008, Efron, 2016) – We deal with network settings.

- Panel data (Arellano and B., 2012) – We allow for dependence across units.

- Variance components in linear regressions on network data (Andrews *et al.*, 2008, Kline *et al.*, 2020) – We estimate other distributional features.

- Empirical Bayes shrinkage (Efron and Morris, 1973, Chen, 2023) – We focus on the marginal density of effects without restricting the conditional density.

Outline

- Model of heterogeneous effects on a network
- Nonparametric identification and estimation
- Estimation under low signal-to-noise
- Applications (preliminary)

Model of heterogeneous effects on a network

Regression model and estimates

- Random coefficients model (abstracting from other covariates):

$$y_i = z_i' \boldsymbol{\eta} + u_i, \quad i = 1, ..., n,$$

that is, in vector form,

$$\boldsymbol{Y} = \boldsymbol{Z}\boldsymbol{\eta} + \boldsymbol{U}.$$

- OLS estimates are

$$\widehat{\boldsymbol{\eta}} = \left(\boldsymbol{Z}'\boldsymbol{Z}\right)^{-1} \boldsymbol{Z}'\boldsymbol{Y}.$$

- We write

$$\widehat{\boldsymbol{\eta}} = \boldsymbol{\eta} + \boldsymbol{V},$$

for $\boldsymbol{V} = \left(\boldsymbol{Z}'\boldsymbol{Z}\right)^{-1} \boldsymbol{Z}'\boldsymbol{U}$.

Assumptions (I): errors

• We assume

$$\boldsymbol{V} \mid \boldsymbol{Z}, \boldsymbol{\eta} \sim \mathcal{N}\left(\mathbf{0}, \boldsymbol{\Sigma}(\boldsymbol{Z})\right).$$

• $\widehat{\eta}_j$'s are approximately normal under suitable conditions, with standard errors σ_j for $j = 1, ..., m$.

• Assuming exact normality is still restrictive. The assumption can be relaxed to some extent [not today].

• The assumption that $\boldsymbol{V}$ is conditionally independent of $\boldsymbol{\eta}$ is not without loss of generality. However, it seems difficult to relax.

• Here we assume $\boldsymbol{\Sigma}(\boldsymbol{Z})$ is known. In practice, we estimate a parameterized version of it (as in Arellano and B., 2012).

Assumptions (II): effects

• Let $\boldsymbol{f}$ denote the joint density of $\eta_1, ..., \eta_m$. Let f_j denote the corresponding marginal density of η_j.

• Our goal is to estimate

$$\bar{f} = \frac{1}{m} \sum_{j=1}^{m} f_j.$$

• Importantly, we do not restrict the conditional density of

$$\eta_1, ..., \eta_m \,\big|\, z_1, ..., z_n.$$

• In particular: (1) η_{j_1} and η_{j_2} may be correlated given $\boldsymbol{Z}$, and (2) $\boldsymbol{\eta}$ may correlate with $\boldsymbol{Z}$.

Economic decisions lead to complex forms of dependence

• The form of $\eta_1, ..., \eta_m \,|\, z_1, ..., z_n$ is governed by a network formation model with heterogeneity.

• Example: Burdett and Mortensen's (1998) model with ex-ante firm heterogeneity in productivity (following Mortensen, 2003):

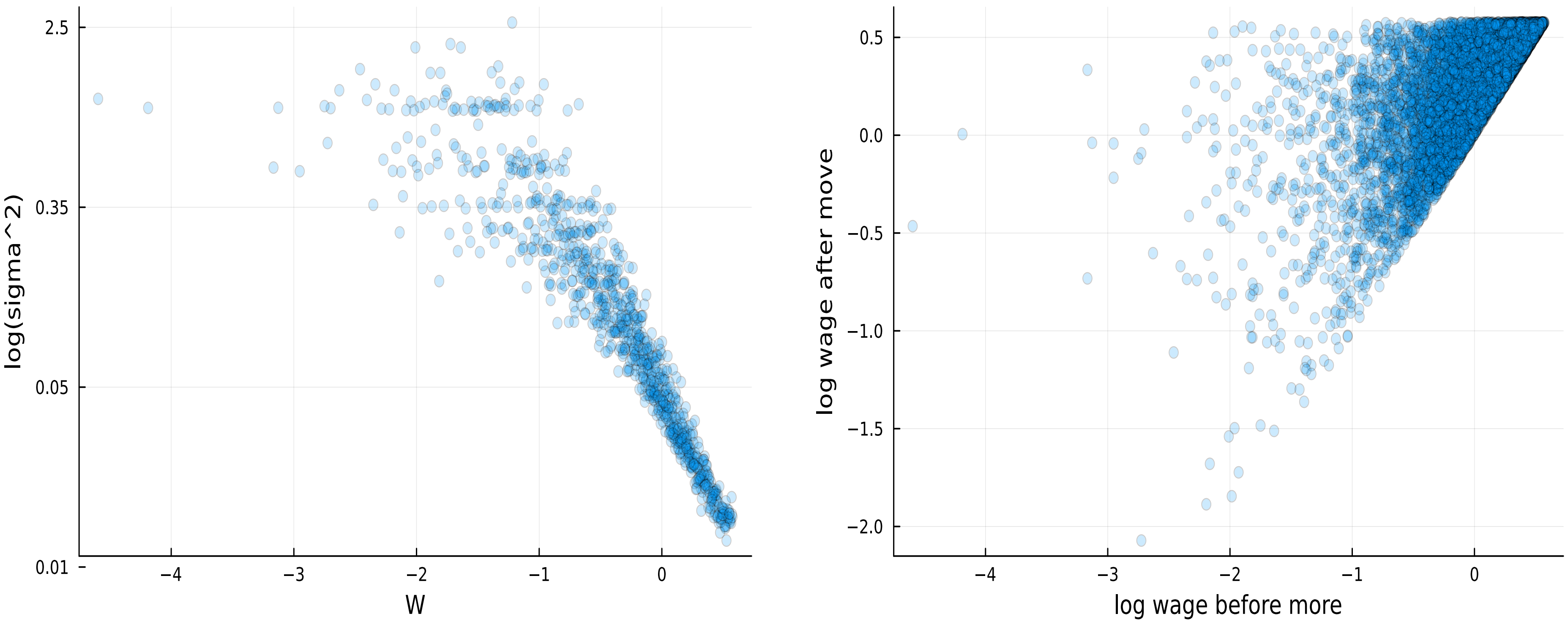
-Wages are increasing in firm productivity.

-On-the-job search leads to correlation between the firm productivities before and after the move.

-High-productivity firms are larger and have more movers.

Numerical illustration based on the Burdett-Mortensen model

(a) wages versus σ_j's (b) wages before and after move



Nonparametric identification and estimation

Characteristic function

- Let $\boldsymbol{i} = \sqrt{-1}$. The characteristic function of η_j is

$$\psi_j(\tau) = \mathbb{E}\left[\exp\left(\boldsymbol{i}\tau\eta_j\right)\right].$$

- The average characteristic function of the η_j's is

$$\psi(\tau) = \frac{1}{m}\sum_{j=1}^{m} \mathbb{E}\left[\exp\left(\boldsymbol{i}\tau\eta_j\right)\right].$$

- Let σ_j be the square root of the jth diagonal element of $\boldsymbol{\Sigma}(\boldsymbol{Z})$; i.e., the standard error of $\widehat{\eta}_j$.

- We have

$$\psi(\tau) = \mathbb{E}\left[\frac{1}{m}\sum_{j=1}^{m} \frac{\exp\left(\boldsymbol{i}\tau\widehat{\eta}_j\right)}{\exp\left(-\frac{1}{2}\sigma_j^2\tau^2\right)}\right].$$

Moments

- Let

$$m_q = \frac{1}{m} \sum_{j=1}^{m} \mathbb{E}\left[{\eta_j}^q\right].$$

- We have

$$m_1 = \mathbb{E}\left[\frac{1}{m} \sum_{j=1}^{m} \widehat{\eta}_j\right],$$

$$m_2 = \mathbb{E}\left[\frac{1}{m} \sum_{j=1}^{m} \left({\widehat{\eta}_j}^2 - \sigma_j^2\right)\right].$$

- More generally, letting H_q be the q-th order ("probabilistic") Hermite polynomial, we find

$$m_q = \mathbb{E}\left[\frac{1}{m} \sum_{j=1}^{m} \sigma_j^q H_q\left(\frac{\widehat{\eta}_j}{\sigma_j}\right)\right].$$

Densities and other distributional features

• For $\overline{f}(\eta)$ the average density of the η_j's, we have, by inverse Fourier transformation,

$$\overline{f}(\eta) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \psi(\tau) \exp\left(-\boldsymbol{i}\tau\eta\right) d\tau.$$

• The corresponding average cdf is (as in Dattner *et al.*, 2011),

$$\overline{F}(\eta) = \frac{1}{2} - \frac{1}{\pi} \int_{0}^{\infty} \frac{1}{\tau} \mathcal{I}\left[\psi(\tau) \exp\left(-\boldsymbol{i}\tau\eta\right)\right] d\tau.$$

• More generally, any nonlinear moment $w_H = \frac{1}{m}\sum_{j=1}^{m} \mathbb{E}[H(\eta_j)]$ satisfies

$$w_H = \frac{1}{2\pi} \int_{-\infty}^{\infty} \psi(\tau)\mathcal{F}[H](-\tau)d\tau,$$

where $\mathcal{F}[H]$ is the Fourier transform of H.

Moment, characteristic function, and density estimates

- An unbiased estimate of $m_q = \frac{1}{m}\sum_{j=1}^{m} \mathbb{E}\left[{\eta_j}^q\right]$ is

$$\widehat{m}_q = \frac{1}{m}\sum_{j=1}^{m} \sigma_j^q H_q\left(\frac{\widehat{\eta}_j}{\sigma_j}\right).$$

- An unbiased estimate of $\psi(\tau) = \frac{1}{m}\sum_{j=1}^{m} \mathbb{E}\left[\exp\left(\boldsymbol{i}\tau\eta_j\right)\right]$ is

$$\widehat{\psi}(\tau) = \frac{1}{m}\sum_{j=1}^{m} \frac{\exp\left(\boldsymbol{i}\tau\widehat{\eta}_j\right)}{\exp\left(-\frac{1}{2}\sigma_j^2\tau^2\right)}.$$

- Let B_m tend to infinity with m (at a suitable rate). An approximately unbiased estimator of the average density $\overline{f}(\eta)$ of the η_j's is

$$\widehat{f}(\eta) = \frac{1}{2\pi}\int_{-B_m}^{B_m} \frac{1}{m}\sum_{j=1}^{m} \frac{\exp\left(\boldsymbol{i}\tau\widehat{\eta}_j\right)}{\exp\left(-\frac{1}{2}\sigma_j^2\tau^2\right)} \exp\left(-\boldsymbol{i}\tau\eta\right) d\tau.$$

Slow convergence rate of nonparametric density estimator

• Assume that ("ordinary smooth" regime in Fan, 1991):

$$|\psi(\tau)| \leq C|\tau|^{-\beta}, \text{ where } \beta > \frac{1}{2}.$$

• Assuming independence across j, and supposing γ known, we can show

$$\mathbb{E}\left[\int \left(\widehat{f}(\eta) - \overline{f}(\eta)\right)^2 d\eta\right] = O\left(B_m^{1-2\beta} + \frac{1}{m} B_m \left\{\max_{j=1,\ldots,m} \mathbb{E}\left[\exp\left(\sigma_j^2 B_m^2\right)\right]\right\}\right).$$

1. Deconvolution with normal errors $\Rightarrow$ the optimal rate is <u>logarithmic</u> in m.

2. The rate is affected by the <u>error variance</u>, which can be large in network settings.

Slow convergence rate (cont.)

• Suppose observations are independent across j. The variance of $\widehat{\psi}(\tau)$ is then

$$\mathrm{Var}\left[\widehat{\psi}(\tau)\right] = \frac{1}{m^2}\sum_{j=1}^{m}\mathbb{E}\left[\exp\left(\sigma_j^2\tau^2\right)\right] - \frac{1}{m}\left|\psi(\tau)\right|^2.$$

• The variance is large when the noise variances σ_j^2 are high.

• Dependence between observations further increases the variance.

• One possibility is to restrict the conditional density of $\boldsymbol{\eta} \mid \boldsymbol{Z}$, in the spirit of Delaigle and Meister (2008) or Chen (2023).

• However, we want to avoid restricting this conditional density, since such restrictions are at odds with economic models.

Estimation under low signal to noise ratio

Parametric marginal specification

• We model the average density $\overline{f}$ of $\eta_1, ..., \eta_m$ as a parametric function $f_{\boldsymbol{\theta}}(\eta)$ of a parameter vector $\boldsymbol{\theta}$.

• Log-spline modeling of Efron (2016):

$$f_{\boldsymbol{\theta}}(\eta) = \frac{\exp(P(\eta)'\boldsymbol{\theta})}{\int \exp(P(\tilde{\eta})'\boldsymbol{\theta}) d\tilde{\eta}},$$

where $P(\eta)$ are polynomial splines.

• Flexible family, leading to a log-concave likelihood function.

• However, since we do not model the conditional density of $\boldsymbol{\eta} \mid \boldsymbol{Z}$, but only the average density $\overline{f}$, we cannot use Efron's estimation method directly.

A simple approach: MLE

- Suppose for simplicity that $\eta_j \sim i.i.d.\, f_{\boldsymbol{\theta}}$ (but do not further restrict the density of $\boldsymbol{\eta}$ given $\boldsymbol{Z}$).

- Start with

$$\widehat{\boldsymbol{\eta}} = \boldsymbol{\eta} + \boldsymbol{V}, \quad \boldsymbol{V} \mid \boldsymbol{Z}, \boldsymbol{\eta} \sim \mathcal{N}\left(\mathbf{0}, \boldsymbol{\Sigma}(\boldsymbol{Z})\right).$$

- Let λ_{max} be the maximal eigenvalue of $\boldsymbol{\Sigma}(\boldsymbol{Z})$, and let

$$\boldsymbol{\varepsilon} \mid \boldsymbol{\eta}, \boldsymbol{V}, \boldsymbol{Z} \sim \mathcal{N}\left(\mathbf{0}, \lambda_{\text{max}} \boldsymbol{I}_m - \boldsymbol{\Sigma}(\boldsymbol{Z})\right).$$

- Then

$$\widehat{\boldsymbol{\eta}} + \boldsymbol{\varepsilon} = \boldsymbol{\eta} + (\boldsymbol{V} + \boldsymbol{\varepsilon}), \quad (\boldsymbol{V} + \boldsymbol{\varepsilon}) \mid \boldsymbol{Z}, \boldsymbol{\eta} \sim \mathcal{N}\left(\mathbf{0}, \lambda_{\text{max}} \boldsymbol{I}_m\right).$$

- This is a textbook deconvolution model with independent normal errors!

- We can use Efron's MLE method to estimate $\boldsymbol{\theta}$.

An alternative approach: QMLE

- An issue with the simple approach is that $\lambda_{\max}$ may be large, especially when the network is poorly connected (Jochmans and Weidner, 2019).

- Let $\sigma^2_{\max}$ denote the maximal diagonal element of $\boldsymbol{\Sigma}(\boldsymbol{Z})$, and let

$$\widetilde{\varepsilon} \mid \boldsymbol{\eta}, \boldsymbol{V}, \boldsymbol{Z} \sim \mathcal{N}\left(\mathbf{0}, \sigma^2_{\max}\boldsymbol{I}_m - \mathrm{diag}\left(\boldsymbol{\Sigma}(\boldsymbol{Z})\right)\right).$$

- Then

$$\widehat{\boldsymbol{\eta}}+\widetilde{\varepsilon} = \boldsymbol{\eta}+(\boldsymbol{V}+\widetilde{\varepsilon}), \; (\boldsymbol{V}+\widetilde{\varepsilon}) \mid \boldsymbol{Z}, \boldsymbol{\eta} \sim \mathcal{N}\left(\mathbf{0}, \boldsymbol{\Sigma}(\boldsymbol{Z}) + \sigma^2_{\max}\boldsymbol{I}_m - \mathrm{diag}\left(\boldsymbol{\Sigma}(\boldsymbol{Z})\right)\right).$$

- In this case, the elements of $(\boldsymbol{V}+\widetilde{\varepsilon})$ are not independent of each other, and they are not independent of $\boldsymbol{\eta}$.

QMLE (cont.)

- We propose to estimate $\boldsymbol{\theta}$ using quasi-MLE, based on the model

$$\widehat{\eta}_j + \widetilde{\varepsilon}_j = \eta_j + \nu_j,$$

where ν_j are i.i.d. $\mathcal{N}(0, \sigma^2_{\max})$ independent of $\eta_1, ..., \eta_m$, and the $\eta_j{}'s$ are i.i.d. draws from $f_{\boldsymbol{\theta}}$.

- This is *quasi*-MLE since we do not model the dependence between the elements $v_j + \widetilde{\varepsilon}_j$ across j.

- Convergence rates for MLE and QMLE:

$$\left\| \widehat{\boldsymbol{\theta}}^{\text{MLE}} - \boldsymbol{\theta} \right\|^2 = O_p\left(\frac{\lambda_{\max}}{m}\right),$$

$$\left\| \widehat{\boldsymbol{\theta}}^{\text{QMLE}} - \boldsymbol{\theta} \right\|^2 = O_p\left(\frac{\sigma^2_{\max}}{m} + \frac{\boldsymbol{\iota}_m' \left(\boldsymbol{\Sigma}(\boldsymbol{Z}) - \boldsymbol{diag}\left(\boldsymbol{\Sigma}(\boldsymbol{Z})\right)\right) \boldsymbol{\iota}_m}{m^2}\right),$$

where $\boldsymbol{\iota}_m$ is a column vector of ones.

Implementation: penalized likelihood

• Following Efron (2016), we add a square penalty to the log-quasi-likelihood and minimize

$$-\sum_{j=1}^{m} \underbrace{\int \ln\left(\int f_{\boldsymbol{\theta}}(\eta) e^{-\frac{1}{2\sigma_{\max}^2}(\widehat{\eta}_j+\varepsilon-\eta)^2} d\eta\right) e^{-\frac{\varepsilon^2}{2\left(\sigma_{\max}^2-\sigma_j^2\right)}} d\varepsilon}_{=L_j(\boldsymbol{\theta})} + \xi\|\boldsymbol{\theta}\|^2.$$

• We discretize the integrals, and use gradient descent for optimization.

• We specify $P(\eta)$ as cubic splines on a set of knots.

• Since the mean of η_j is not identified, we impose the normalization $\sum_{j=1}^{m} \eta_j = 0$ (which is not numerically equivalent to imposing $\eta_1 = 0$, for example).

Implementation: tuning parameters

• We set ξ and the number of knots by maximizing the marginal likelihood (as in Ruppert *et al.*, 2003):

$$\int \exp\left(\sum_{j=1}^{m} L_j(\boldsymbol{\theta})\right) \frac{1}{(2\pi)^{\frac{m}{2}}(2\xi)^{-\frac{m}{2}}} \exp\left(-\xi\|\boldsymbol{\theta}\|^2\right) d\boldsymbol{\theta}.$$

• In practice, we use a Laplace approximation to the integral and maximize

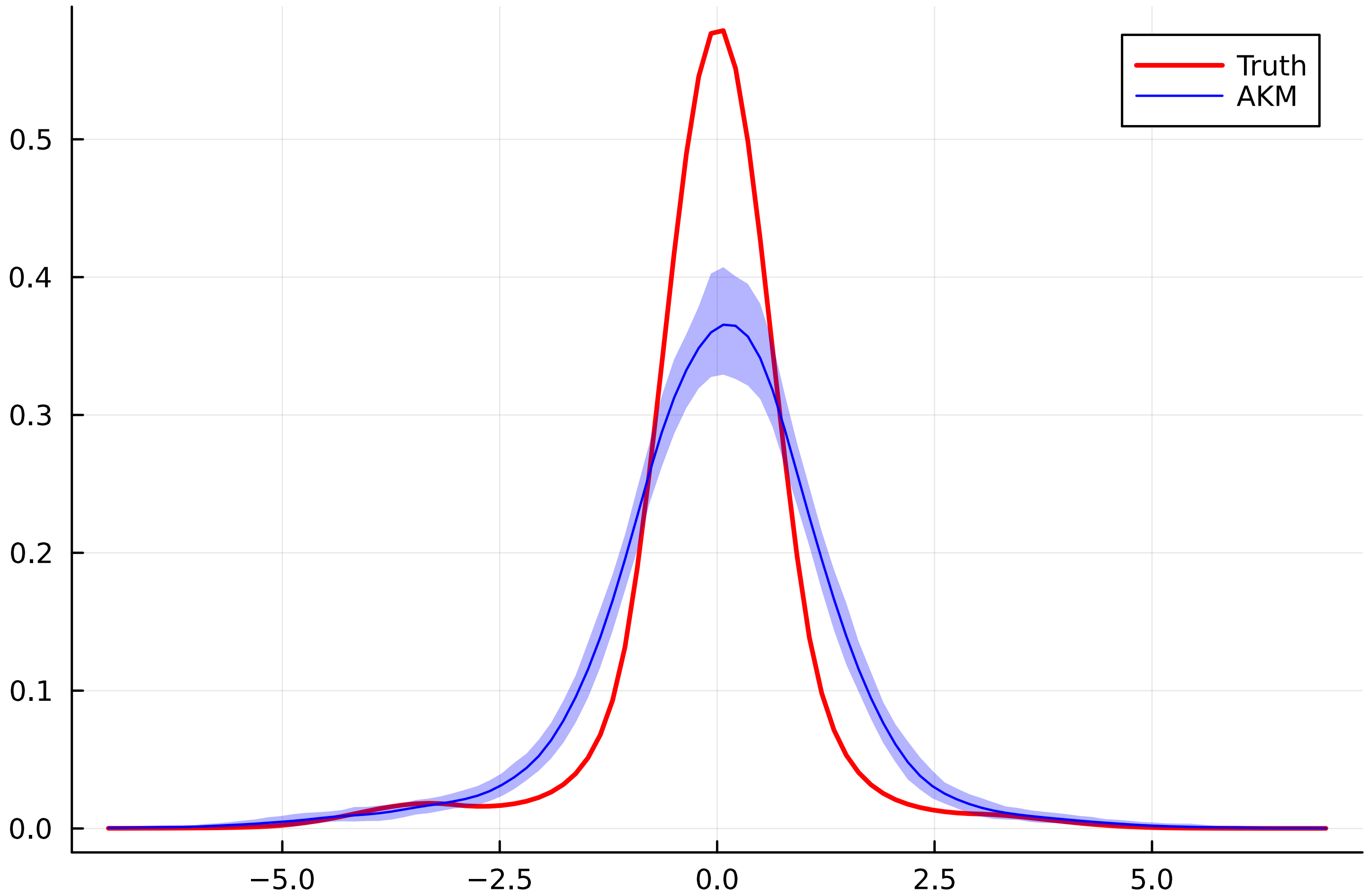
$$\sum_{j=1}^{m} L_j\left(\widehat{\boldsymbol{\theta}}(\xi)\right) - \frac{m}{2}\ln(2\pi) + \frac{m}{2}\ln(2\xi) - \xi\left\|\widehat{\boldsymbol{\theta}}(\xi)\right\|^2 + \frac{1}{2}\ln\det\left(\boldsymbol{H}(\xi)\right),$$

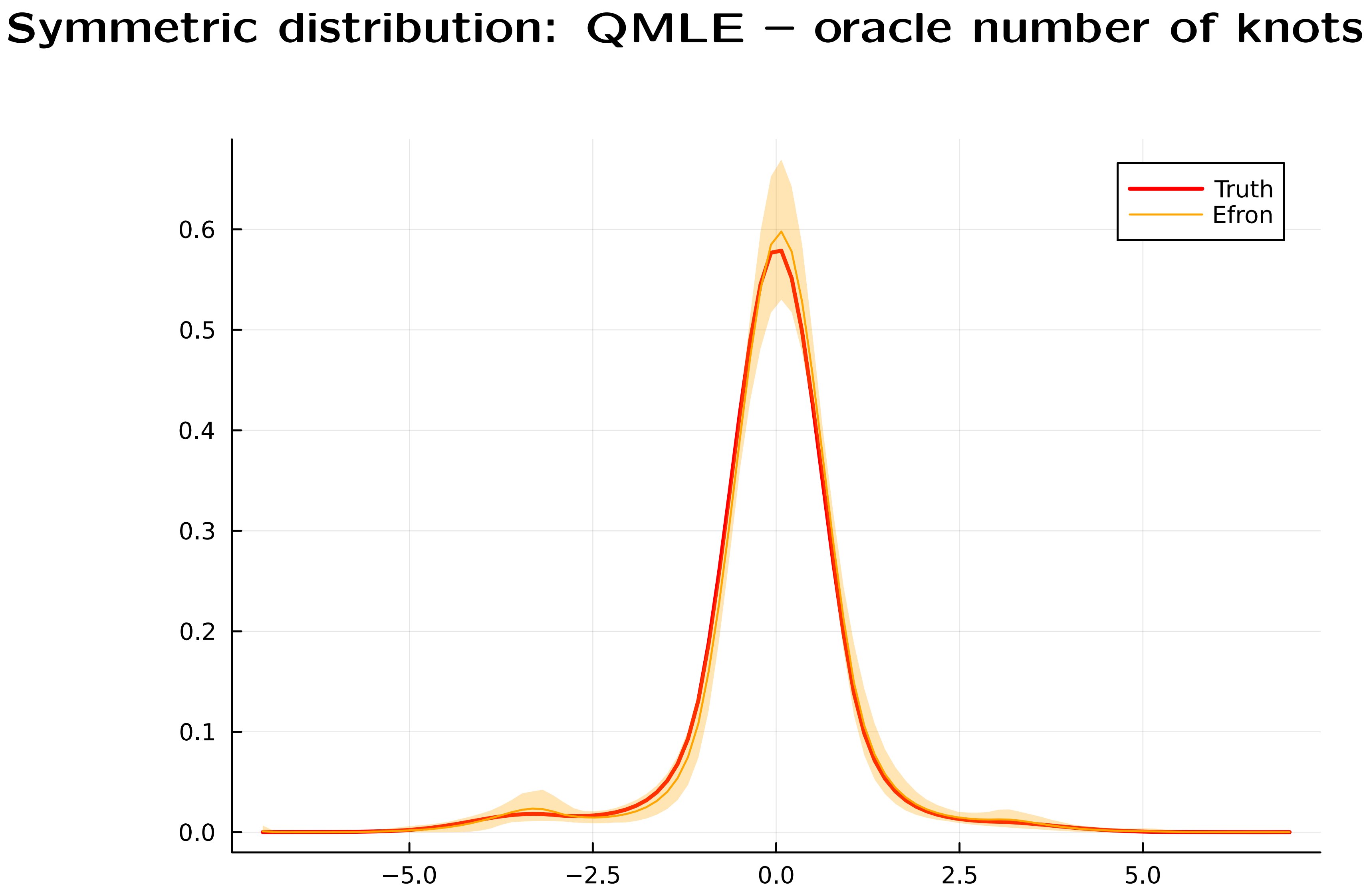
where $\widehat{\boldsymbol{\theta}}(\xi)$ is the QMLE for given ξ, and

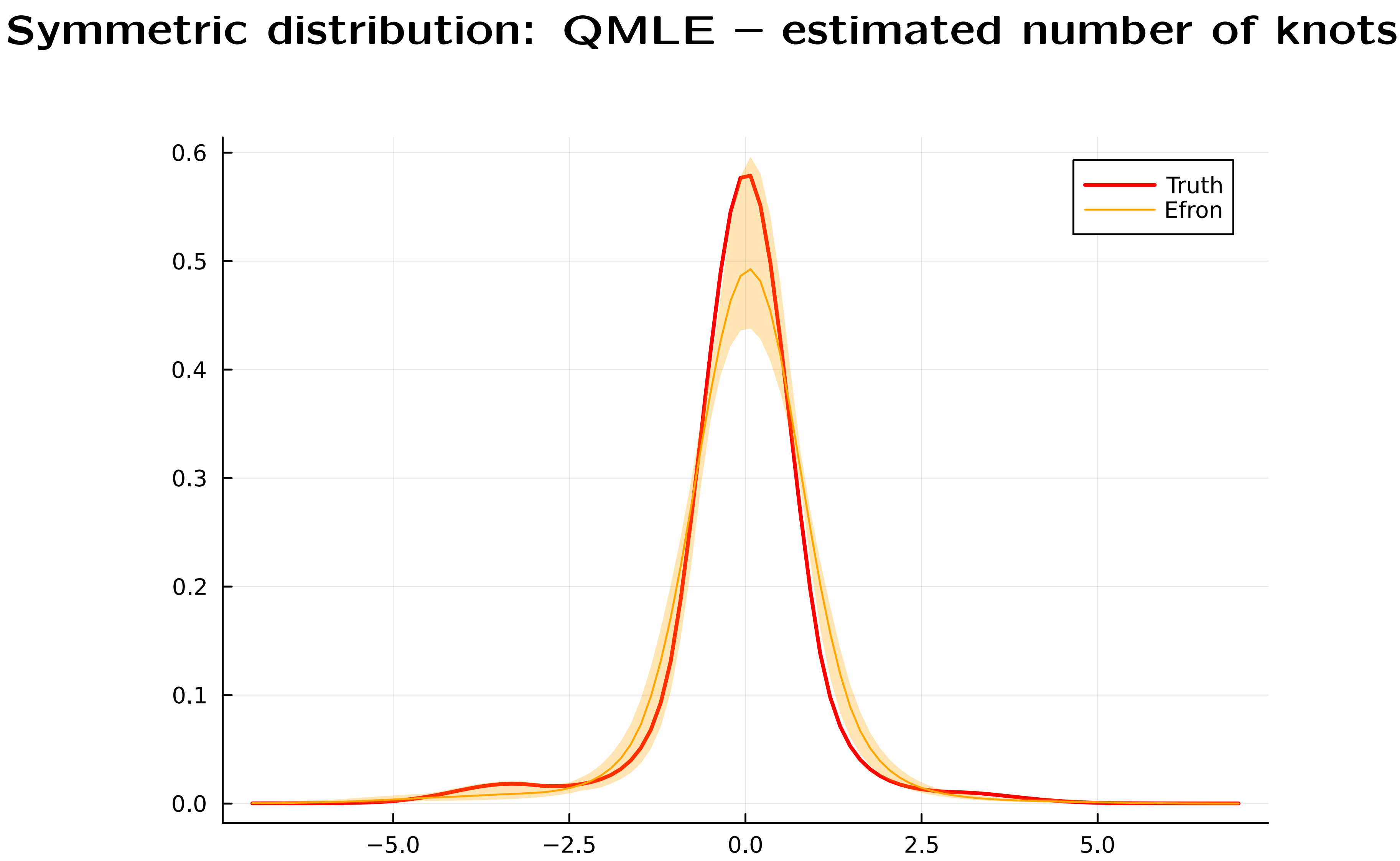
$$\boldsymbol{H}(\xi) = -\frac{\partial^2}{\partial\theta\partial\theta'}\bigg|_{\theta=\widehat{\boldsymbol{\theta}}(\xi)} \left\{\sum_{j=1}^{m} L_j(\theta) - \xi\|\theta\|^2\right\}.$$

Applications (preliminary)

Symmetric distribution: fixed-effects estimates $\widehat{\eta}_j$ (AKM, from Abowd, Kramarz and Margolis 1999)

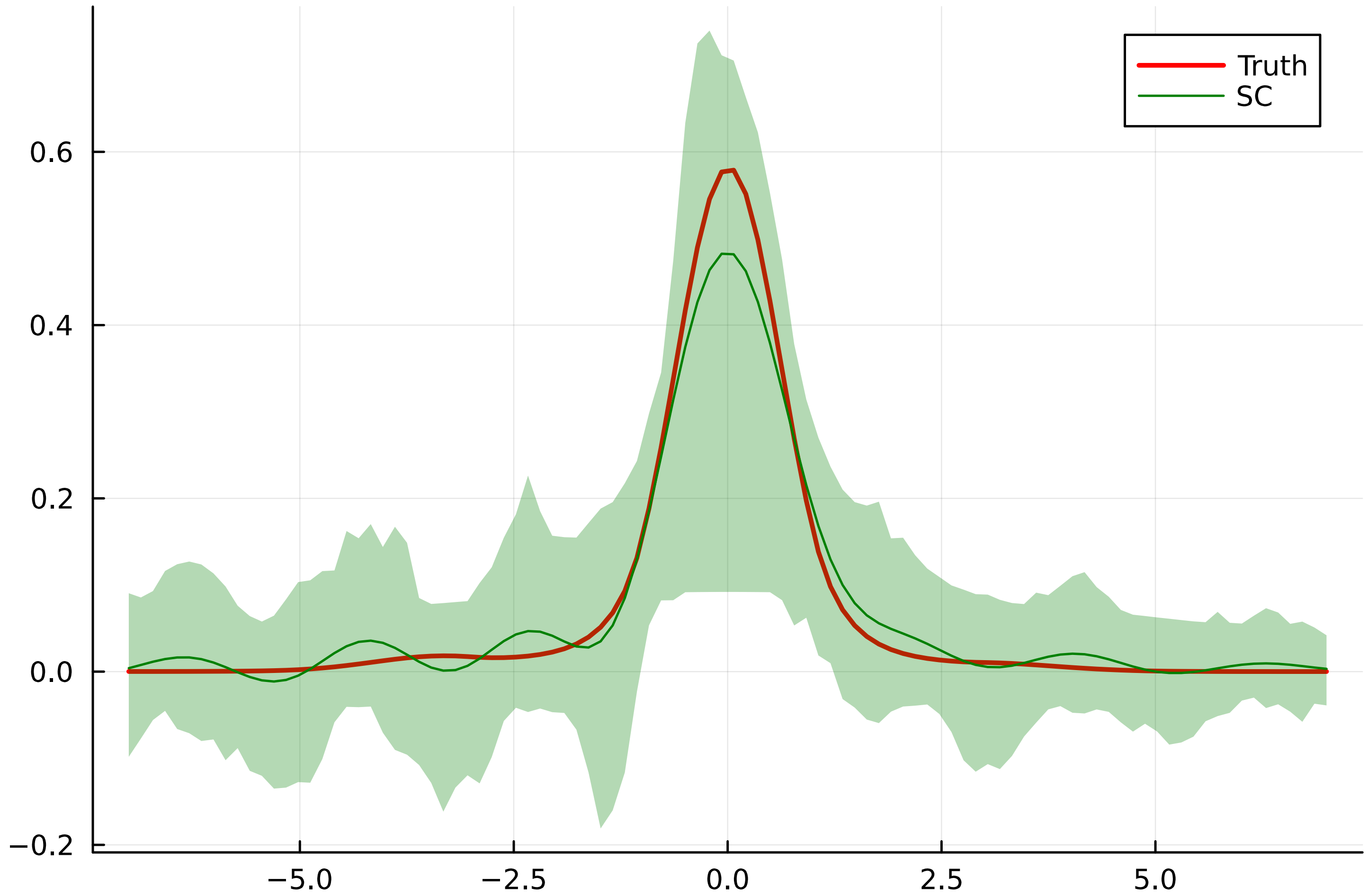



Symmetric distribution: QMLE − oracle number of knots
Truth
Efron
0.0
0.1
0.2
0.3
0.4
0.5
0.6
−5.0
−2.5
0.0
2.5
5.0

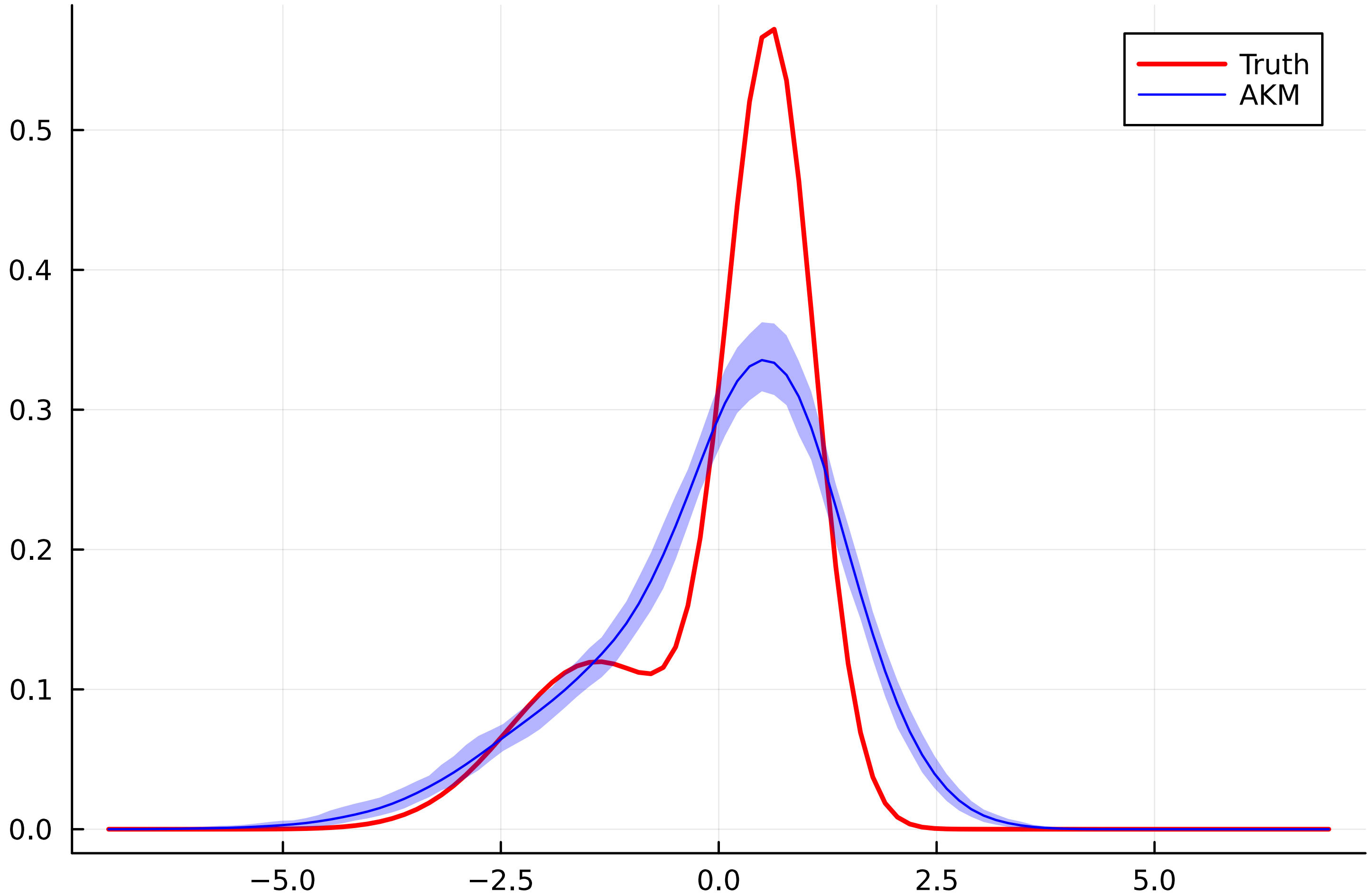


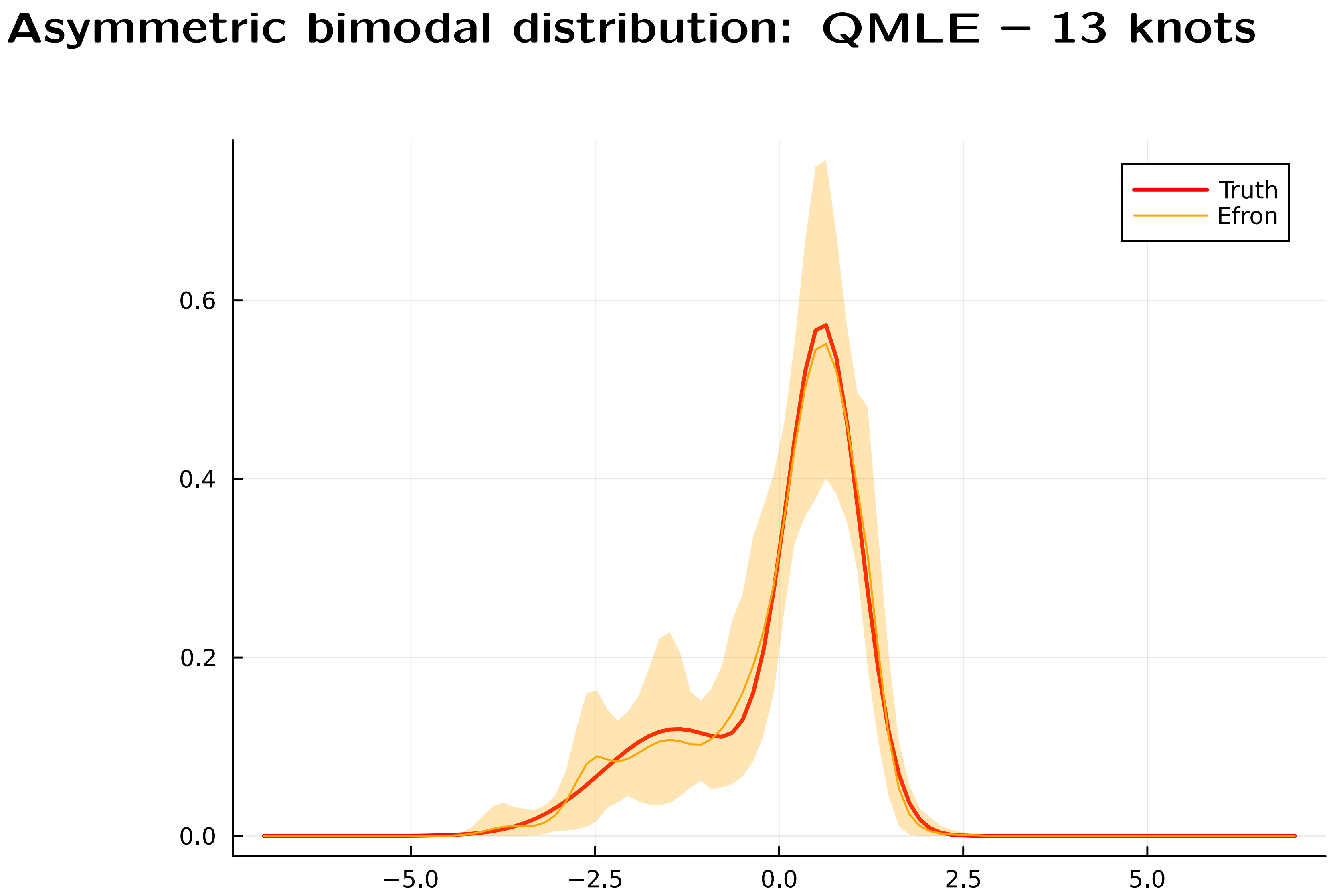
Symmetric distribution: QMLE – estimated number of knots
Truth
Efron
0.6
0.5
0.4
0.3
0.2
0.1
0.0
−5.0
−2.5
0.0
2.5
5.0

Symmetric distribution: nonparametric kernel deconvolution (in the spirit of Stefanski and Carroll, 1990)

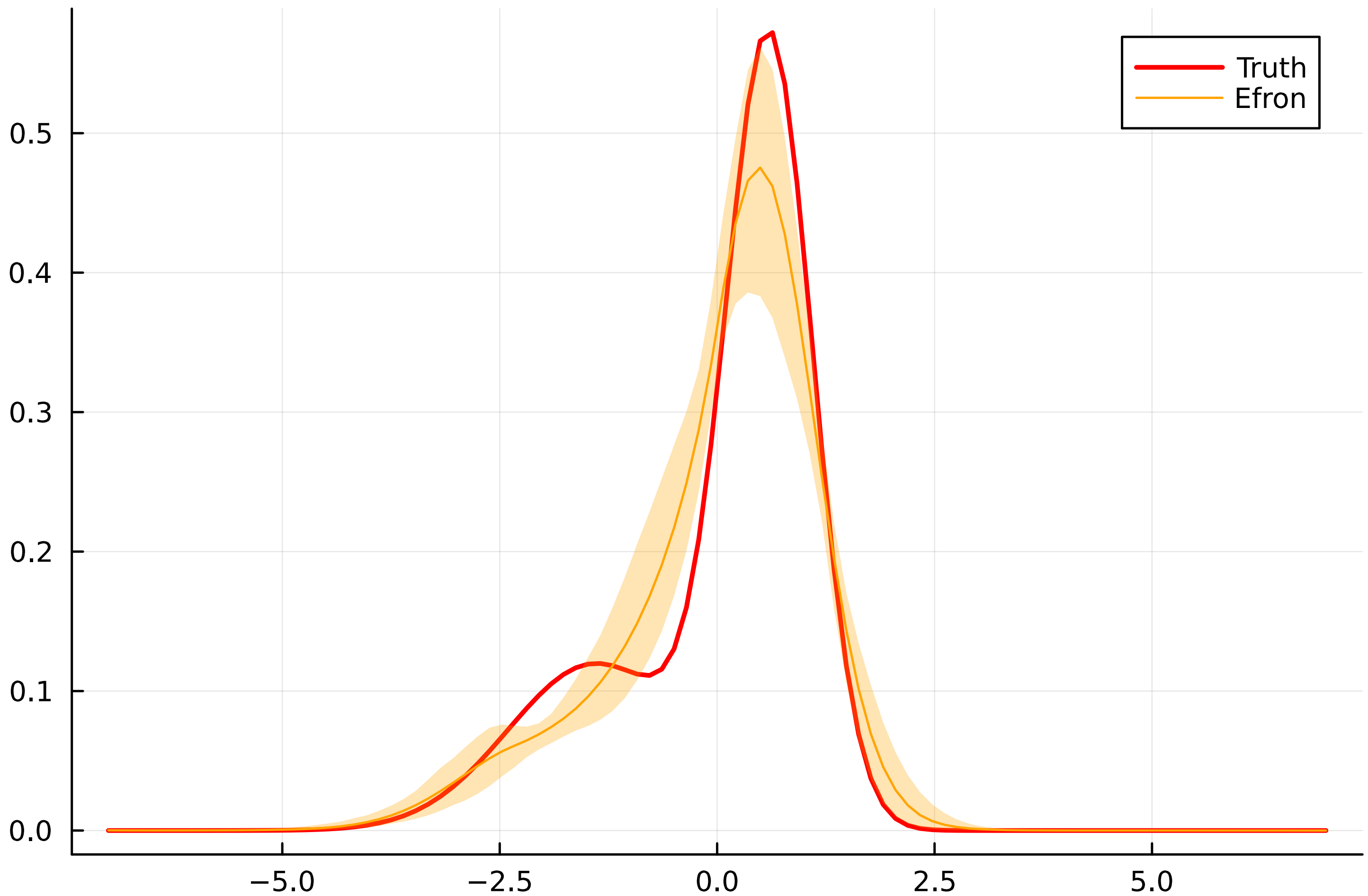


Asymmetric bimodal distribution: fixed-effects estimates $\widehat{\eta}_j$

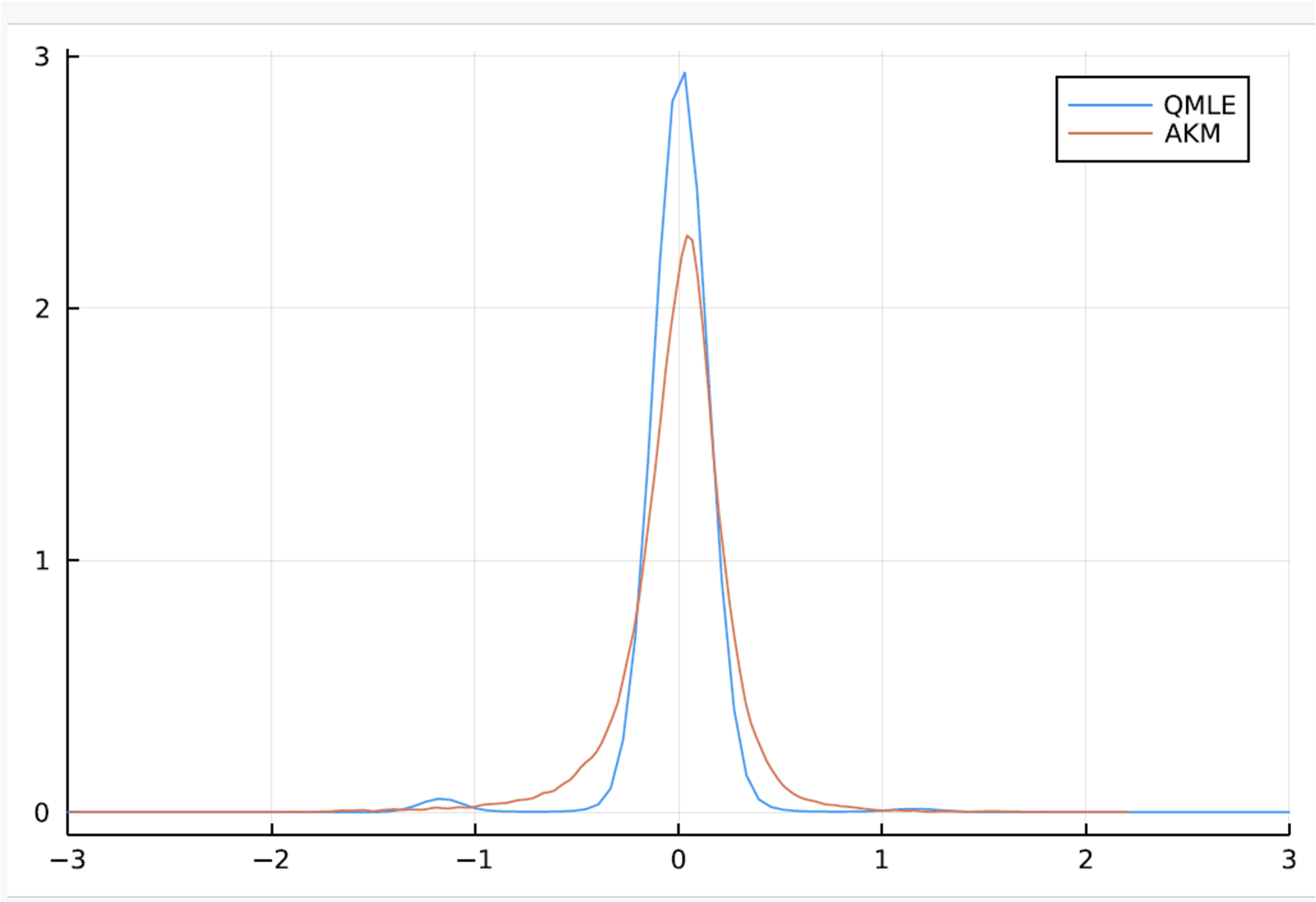



Asymmetric bimodal distribution: QMLE − 13 knots
Truth
Efron
0.0
0.2
0.4
0.6
−5.0
−2.5
0.0
2.5
5.0

Asymmetric bimodal distribution: QMLE – estimated number of knots

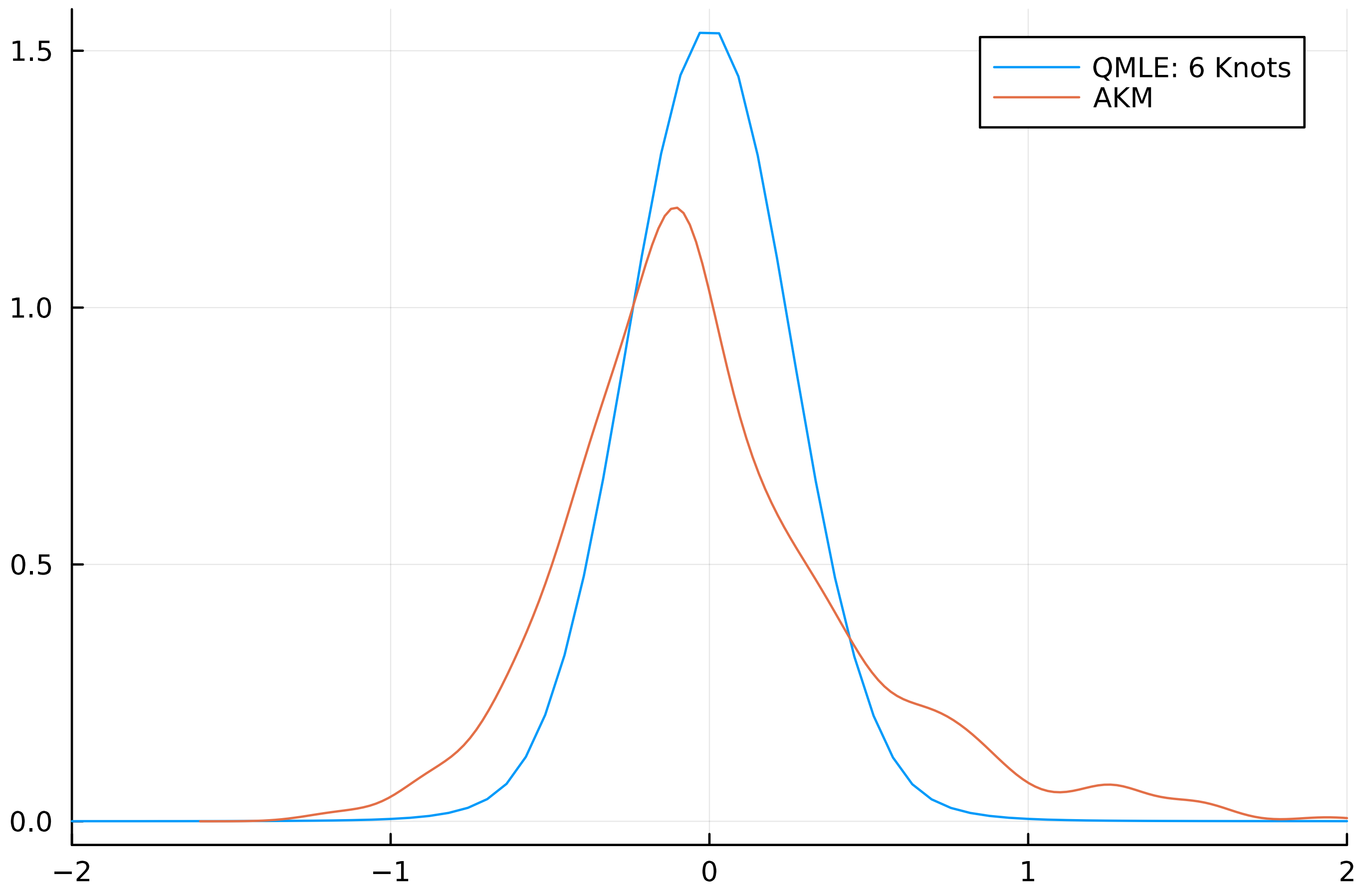


Swedish administrative data: firm effects – 10 knots



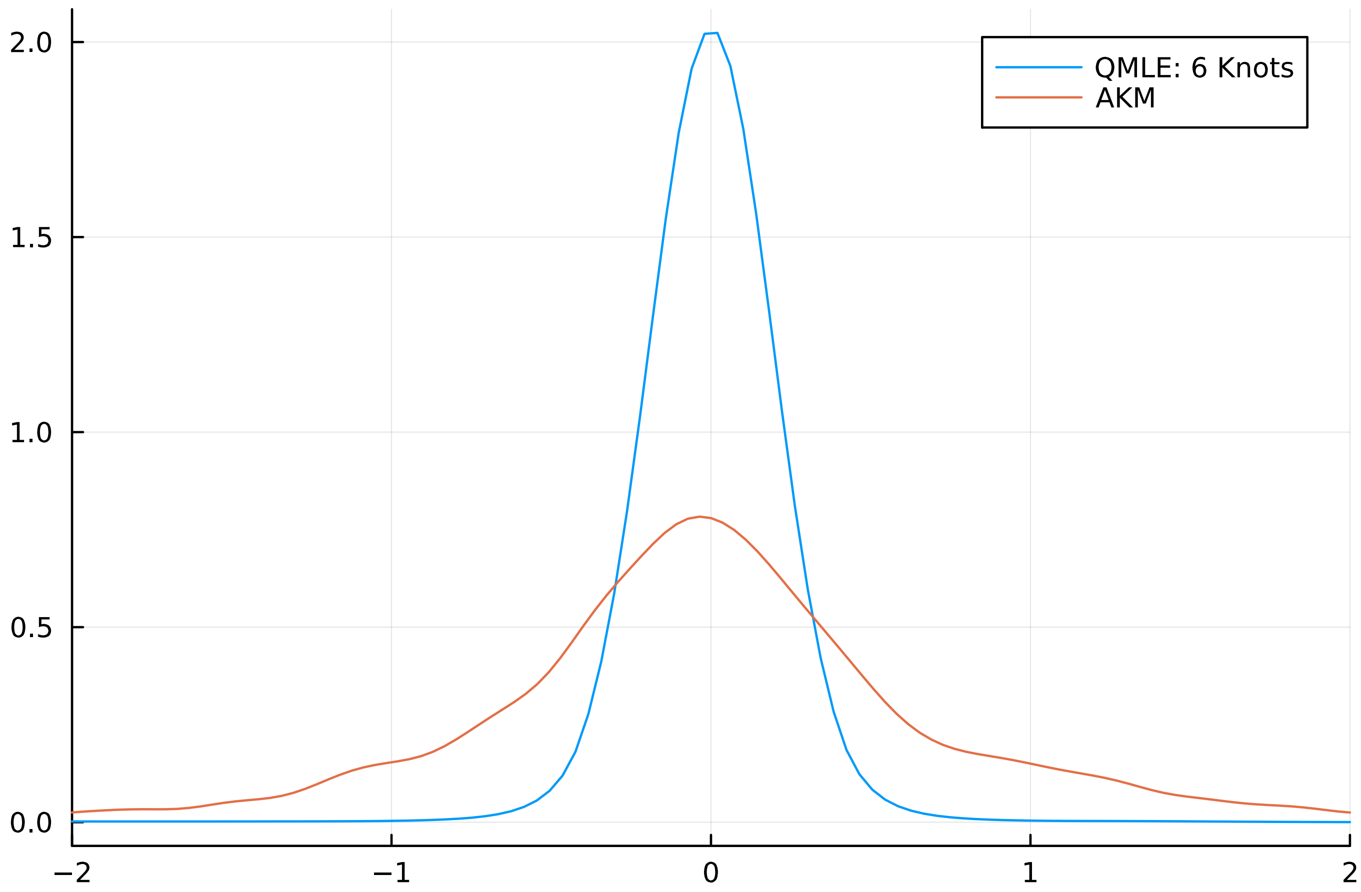
Notes: Sample from B., Lamadon and Manresa (2019).

US administrative data: neighborhood effects (commuting zones, unweighted) – estimated number of knots



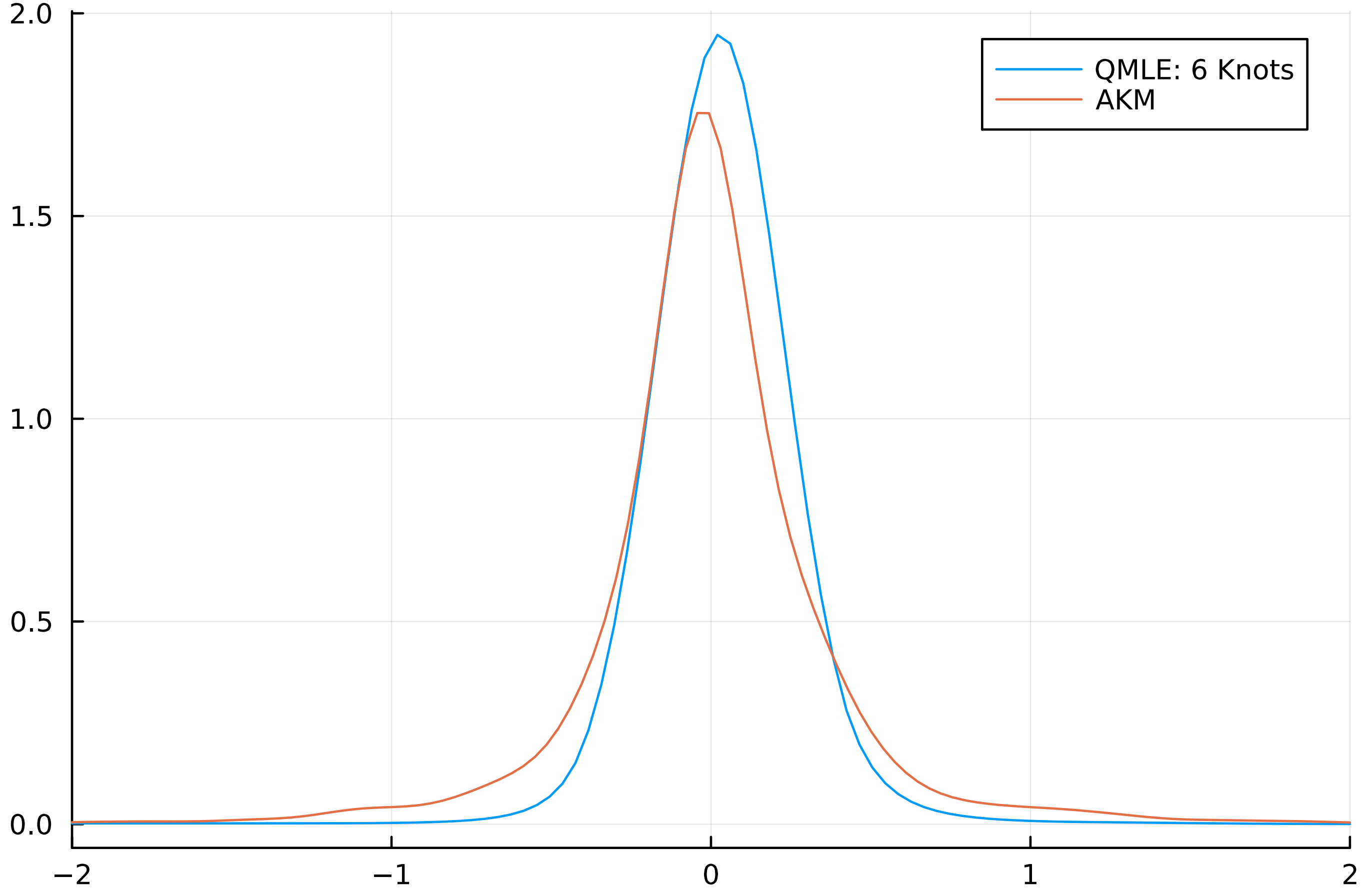
Notes: Commuting zone estimates from Chetty and Hendren (2018).

US administrative data: neighborhood effects (counties, unweighted) – estimated number of knots



Notes: County-level estimates from Chetty and Hendren (2018).

US administrative data: neighborhood effects (counties, weighted by population density) – estimated number of knots



Notes: County-level estimates from Chetty and Hendren (2018).

Extension: bivariate densities and sorting

• We aim at developing practical methods to estimate distributions in network settings.

• In addition to marginal densities of firm effects, we are interested in joint densities of firm and worker effects:

$$\begin{pmatrix} \widehat{\eta}_{\text{worker}} \\ \widehat{\eta}_{\text{firm}} \end{pmatrix} = \begin{pmatrix} \eta_{\text{worker}} \\ \eta_{\text{firm}} \end{pmatrix} + \text{Noise}.$$

• We also want to estimate joint densities of firm effects before and after a job move:

$$\begin{pmatrix} \widehat{\eta}_{\text{before}} \\ \widehat{\eta}_{\text{after}} \end{pmatrix} = \begin{pmatrix} \eta_{\text{before}} \\ \eta_{\text{after}} \end{pmatrix} + \text{Noise}.$$

• Our substantive goal is to document the shapes of inequality and sorting nonparametrically.

Additional slides

Convergence rate of moment estimators

- We have

$$\mathbb{E}\left[(\widehat{m}_2 - m_2)^2\right] = O\left(\frac{\text{Var}(\eta'\eta)}{m^2} + \mathbb{E}\left[\frac{\boldsymbol{\eta}'\boldsymbol{\Sigma}(\boldsymbol{Z})\boldsymbol{\eta}}{m^2}\right] + \mathbb{E}\left[\frac{\text{Tr}\left(\boldsymbol{\Sigma}(\boldsymbol{Z})^2\right)}{m^2}\right]\right).$$

-Related to Kline *et al.* (2020), who study the estimation of quadratic forms in $\boldsymbol{\eta}$.

- More generally, the <u>convergence rate</u> of $\widehat{m}_q$ depends on:

-the dependence across j of powers of (η_j, σ_j).

-the dependence across j of v_j.

Estimating error variances

- In our application, we assume $\mathbb{E}(\boldsymbol{U} \mid \boldsymbol{\eta}, \boldsymbol{Z}) = \mathbf{0}$ and $\text{Var}(\boldsymbol{U} \mid \boldsymbol{Z}) = \boldsymbol{\Omega}(\boldsymbol{Z}; \boldsymbol{\gamma})$, where $\boldsymbol{\gamma}$ is a low-dimensional parameter.

- Specifically, we assume

$$\boldsymbol{\Omega}(\boldsymbol{Z}; \gamma) = \gamma \boldsymbol{I}_n,$$

and rely on

$$\widehat{\gamma} = \frac{\boldsymbol{Y}' \left(\boldsymbol{I}_n - \boldsymbol{Z}(\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}' \right) \boldsymbol{Y}}{n - m},$$

as in Andrews *et al.* (2008).

- Then, the estimates $\widehat{\sigma}_j^2$ are the diagonal elements of

$$\widehat{\boldsymbol{\Sigma}}(\boldsymbol{Z}) = \left(\boldsymbol{Z}'\boldsymbol{Z}\right)^{-1} \boldsymbol{Z}'\boldsymbol{\Omega}(\boldsymbol{Z}; \widehat{\gamma}) \, \boldsymbol{Z} \left(\boldsymbol{Z}'\boldsymbol{Z}\right)^{-1} = \widehat{\gamma} \left(\boldsymbol{Z}'\boldsymbol{Z}\right)^{-1}.$$

Monte Carlo designs

• The DGPs are generated based on 6 parameters: 2 of them control the size of the network, and 4 of them control its structure.

-One key parameter controls the correlation between η_j and σ_j^2.

-Another key parameter controls the correlation between the η_j's before and after a move.

• We vary the shape of $f(\eta)$: normal, symmetric and peaked, asymmetric and bimodal.

• We pick parameters to mimic Swedish matched employer employee data (B. *et al.*, 2019, 2023).