

IZA DP No. 5847

Using the Helmert-Transformation to Reduce Dimensionality in a Mixed Model: Application to a Wage Equation with Worker and Firm Heterogeneity

Øivind A. Nilsen
Arvid Raknerud
Terje Skjerpen

July 2011

Using the Helmert-Transformation to Reduce Dimensionality in a Mixed Model: Application to a Wage Equation with Worker and Firm Heterogeneity

Øivind A. Nilsen

*Norwegian School of Economics
and IZA*

Arvid Raknerud

Statistics Norway

Terje Skjerpen

Statistics Norway

Discussion Paper No. 5847
July 2011

IZA

P.O. Box 7240
53072 Bonn
Germany

Phone: +49-228-3894-0

Fax: +49-228-3894-180

E-mail: iza@iza.org

Any opinions expressed here are those of the author(s) and not those of IZA. Research published in this series may include views on policy, but the institute itself takes no institutional policy positions.

The Institute for the Study of Labor (IZA) in Bonn is a local and virtual international research center and a place of communication between science, politics and business. IZA is an independent nonprofit organization supported by Deutsche Post Foundation. The center is associated with the University of Bonn and offers a stimulating research environment through its international network, workshops and conferences, data service, project support, research visits and doctoral program. IZA engages in (i) original and internationally competitive research in all fields of labor economics, (ii) development of policy concepts, and (iii) dissemination of research results and concepts to the interested public.

IZA Discussion Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be available directly from the author.

ABSTRACT

Using the Helmert-Transformation to Reduce Dimensionality in a Mixed Model: Application to a Wage Equation with Worker and Firm Heterogeneity^{*}

A model for matched data with two types of unobserved heterogeneity is considered – one related to the observation unit, the other to units to which the observation units are matched. One or both of the unobserved components are assumed to be random. This mixed model allows identification of the effect of time-invariant variables on the observation units. Applying the Helmert transformation to reduce dimensionality simplifies the computational problem substantially. The framework has many potential applications; we apply it to wage modeling. Using Norwegian manufacturing data shows that the assumption with respect to the two types of heterogeneity affects the estimate of the return to education considerably.

JEL Classification: C23, C81, J31

Keywords: high-dimensional two-way unobserved components,
matched employer-employee data, ECM-algorithm

Corresponding author:

Arvid Raknerud
Statistics Norway
PO Box 8131 Dep
NO-0033 Oslo
Norway
E-mail: rak@ssb.no

^{*} Comments and suggestions made during the presentation at the Nordic Econometric Meeting 2011, and on a previous version of this paper at ESEM 2009 in Barcelona are greatly appreciated.

1 Introduction

Access to matched data sets enables consideration of unobserved heterogeneity corresponding to different types of units in regression analyses. Often the main focus is on one type of observational unit, but it is also necessary to account for unobserved heterogeneity caused by another type of observational unit that is matched to the main type. Wage modeling by means of matched employer–employee data, which is the topic of the current paper, may be the best-known example. Here, the individual is considered the main observational unit, whereas the firm to which the individual is matched has the role of a secondary observational unit. For consistent and efficient estimation of the effects of observed explanatory variables it is vital to account for both individual- and firm-specific unobserved heterogeneity. Using only individual-level data may yield misleading policy implications.

However, other fields in economics may have a corresponding data design. Let us mention three examples, which we do not claim are exhaustive. If there are matched data for banks and their customers, one may account for both unobserved bank and bank customer-specific effects.¹ A second example could be connected to FDI. A domestic firm is matched to a foreign country, and it is desirable to account for unobserved heterogeneity stemming both from the firm itself and from the country in which the firm is involved.² A final example is taken from health economics in a modeling framework where the main observational unit consists of patients and where they are matched to general practitioners. With such data, unobserved heterogeneity related both to the patients and to the general practitioners may be considered.³

Returning to wage modeling, Abowd *et al.* (1999), whose paper constitutes a

¹For instance, this is the case in Ioannidou and Ongena (2010).

²For an overview of analyses of FDI in a panel data context, see Blanchard *et al.* (2008).

³For a panel data analysis employing matched data of this type, see for instance Godager and Biørn (2010).

seminal contribution with respect to wage modeling using employer–employee data, represented both unobserved individual- and firm-specific heterogeneity by fixed effects. In applications, the researcher is often interested in the effect of observed time-invariant variables, or of variables that may almost be regarded as such. An example is the length of education, which for most individuals does not vary over the sample period. However, the fixed effects specification has the problematic feature that one cannot identify the effects of variables that are constant over time. For example, the effect of a change in education is identified when the individual effects are random, but not when they are fixed.⁴ Another advantage of the random components model is that it is far more parsimonious with respect to the number of parameters than the fixed effects model.

In this paper, we consider a linear mixed model with an unobserved effect corresponding to the main observation unit (e.g., an individual) and an unobserved effect corresponding to another type of unit (e.g., a firm) with which the main observation unit is matched at a given point in time.⁵ The matching between the two types of units may change over time, and is considered to be the outcome of an exogenous matching variable. We allow the unobserved effects corresponding to the matched units to be correlated. Before estimating the parameters of the regression equation we apply the Helmert transformation to reduce the dimensionality problem associated with a possibly very large number of latent variables.⁶ The main contribution of this paper is to show that, within a random effects framework, the Helmert trans-

⁴There may also be intermediate cases in a situation with several covariates when it is possible to identify the effect of one-dimensional variables even in the presence of fixed effects. However, this requires an *a priori* assumption stating that some of the covariates are uncorrelated with the random unobserved individual-specific term. For this approach, cf. Hausman and Taylor (1981).

⁵For the statistical treatment of linear mixed models, cf. for instance Searle *et al.* (1992) and Demidenko (2004).

⁶Balestra and Krishnakumar (2008) and Arellano and Bover (2005) comment on this transformation even though they do not use the label Helmert transformation. Rather, they refer to it as the backward and forward orthogonal deviations operator. See Keane and Runkle (1992) for the related concept of forward filtering.

formation can be used to sweep out the random effects corresponding to the main observation unit. The resulting profile likelihood will then have much fewer latent variables than the original model, that is, equal to the number of main units plus the number of units to which these units can be matched. To estimate the parameters of the models we propose an Expectation Conditional Maximization (ECM) algorithm (see Meng and Rubin, 1993) to maximize the profile log-likelihood function.

In an application, we investigate the best specification of unobserved heterogeneity in a wage equation when there is access to unbalanced employer–employee panel data. What one ultimately seeks is a test corresponding to the standard Hausman test applied in panel data models where one only addresses one-way unobserved heterogeneity. Models that include random individual and firm effects as well as random individual and fixed firm effects are of substantial interest—both types of model allow for the identification of the effects of time-invariant individual-specific variables, but the latter specification is less restrictive.

We apply our modeling framework to a sample of individuals working in a traditional Norwegian manufacturing industry, production of machinery (NACE 29). Panel employer–employee data for the years 1995–2006 are used. The final data consist of 15,415 observations. We have 2,021 individuals and 770 firms. As observed individual covariates in the wage equation, we use length of education, a third-order polynomial in experience, three dummies for type of education, a dummy for gender, five dummies for labor market areas and 11 year dummies. Of the skill-related variables, only those involving experience vary across both individuals and time.

Specifying both the unobserved individual-specific effects and the firm-specific effects as random effects, we find the coefficient of years of education to differ only modestly from the estimate in the model with individual random effects and no firm effects. This is not very surprising. If the first specification is valid, we know that the

covariance matrix of the gross error term, which is a sum of two one-dimensional random terms and a genuine error term, will have a certain structure and that this error term is independent of the explanatory variables. Accounting for this structure is necessary to obtain efficient estimates of the slope parameters of the wage equation, but not to obtain consistent estimates of these parameters. Furthermore, the estimate of the correlation coefficient between the individual- and firm-specific random effects is statistically different from zero. Constraining this parameter to zero does not produce estimated slope parameters that are very different from those obtained in the specification where it is allowed to be estimated as a free parameter. However, as emphasized by Eeckhout and Kircher (2010), it is not straightforward to interpret such an empirical finding. In contrast, the model specification with random individual effects and fixed firm effects does produce substantially different estimates with regard to returns to education. This model is more flexible than the model where both the individual and firm effects are random, because the fixed firm effects are not constrained to be independent of the explanatory variables, which may explain the difference in the parameter estimates.

The rest of the paper is organized as follows. In Section 2, we outline the general modeling framework, introduce the Helmert transformation and present the estimation algorithm. Section 3 contains an application on wage equation estimation and discusses various specifications. Some concluding remarks are provided in Section 4.

2 The general model

The starting point of our analysis is the following model with a three-way structure:

$$y_{ijt} = x_{it}\beta + z_i\gamma + \mu_i + \nu_j + \eta_{ijt}, \quad (1)$$

where y_{ijt} is the endogenous variable for observation unit i , matched with unit j , and observed at time t . In matched employer–employee data, j will typically denote the firm or employer of individual i at t , but other applications are obviously possible; for example, i may denote a firm and j its (main) bank (see Ioannidou and Ongena, 2010). For specificity, we henceforth refer to i as an individual and j as a firm. Then x_{it} represents the time-varying covariates of individual i , z_i represents the time-invariant covariates, μ_i is a random effect corresponding to individual i (henceforth “individual effect”), ν_j is a random effect corresponding to firm j (“firm effect”) and η_{ijt} is a genuine error term.

The index j is assumed to be the outcome of a stochastic index function $j = J(i, t) \in \{1, 2, \dots, M\}$, denoting the unit matched to i at t . We assume throughout that the distribution of η_{ijt} does not depend on j . Then we can drop the subscript j from y_{ijt} and η_{ijt} , and rewrite (1) as follows:

$$y_{it} = x_{it}\beta + z_i\gamma + \mu_i + \nu_{J(i,t)} + \eta_{it}, \quad (2)$$

where $E(\eta_{it}) = 0$ and $E(\eta_{it}^2) = \sigma_{\eta\eta}$ for all i, t . Letting $\nu = (\nu_1, \dots, \nu_M)'$ denote the vector of all the M random effects and G_{it} an appropriate selection vector, such that $G_{it}\nu = \nu_{J(i,t)}$, we can write

$$y_{it} = x_{it}\beta + z_i\gamma + \mu_i + G_{it}\nu + \eta_{it}. \quad (3)$$

To simplify the notation, we assume that all individuals enter the sample at $t = 1$ (or, equivalently, we can redefine t to denote the t 'th observation on individual i). We allow for unbalanced data, with unit i exiting the sample at $t = T_i$.

To sweep out the individual effects from models with both individual and firm effects, we propose to use the Helmert transformation. Formally, the Helmert trans-

formation of y_{it} , $t = 1, \dots, T_i$, is given by $(\vec{y}_{i,1}, \dots, \vec{y}_{i,T_i-1}, \bar{y}_i)$, where

$$\vec{y}_{i,t} = \sqrt{t/(t+1)} \left(y_{i,t+1} - t^{-1} \sum_{s=1}^t y_{i,s} \right), \quad t = 1, \dots, T_i - 1,$$

and

$$\bar{y}_i = T_i^{-1} \sum_{s=1}^{T_i} y_{i,s}.$$

A corresponding transformation can also be applied component wise to the variables included in an arbitrary vector, say x . It is easy to check that all the corresponding Helmert transformed error terms, $\vec{\eta}_{i,t}$ and $\bar{\eta}_i$, are uncorrelated, given that the original error terms, η_{it} , are uncorrelated and homoscedastic (i.e., have constant variance over time). Moreover, the individual effects will be swept out from all the transformed variables, except \bar{y}_i . Of course, the Helmert transformation is not the only way of sweeping out the individual effects (see, for example, Andrews *et al.* (2008) for a discussion of the within estimator in this context), but it has the huge advantage of preserving the orthogonality of the error terms.

Fixed individual and firm effects Let us first consider the estimator when both the individual and the firm effects are fixed. The estimator is then obtained by minimizing the quadratic form

$$Q(\beta, \gamma, \nu, \mu) = \sum_{i=1}^N T_i (\bar{y}_i - \bar{x}_i \beta - z_i \gamma - \mu_i - \bar{G}_i \nu)^2 + (\vec{y} - \vec{X} \beta - \vec{G} \nu)' (\vec{y} - \vec{X} \beta - \vec{G} \nu)$$

with respect to $(\beta, \gamma, \nu, \mu)$, where $\mu = (\mu_1, \dots, \mu_N)'$, $\bar{x}_i = T_i^{-1} \sum_{s=1}^{T_i} x_{is}$, $\bar{G}_i = T_i^{-1} \sum_{s=1}^{T_i} G_{is}$,

$$\vec{y} = (\vec{y}'_1, \dots, \vec{y}'_N)'$$

with $\vec{y}'_i = (\vec{y}'_{i,1}, \dots, \vec{y}'_{i,T_i-1})'$, and

$$\vec{G} = (\vec{G}'_1, \dots, \vec{G}'_N)'$$

with $\vec{G}_i = (\vec{G}_{i;1}', \dots, \vec{G}_{i;T_i-1}')$. Note that \vec{G} has dimension $(\sum_{i=1}^N (T_i - 1)) \times M$ and \bar{G}_i dimension $N \times M$.

The first-order conditions for minimizing $Q(\beta, \gamma, \nu, \mu)$ then become

$$\bar{y}_i - \bar{x}_i\beta - z_i\gamma - \mu_i - \bar{G}_i\nu = 0 \quad (4)$$

and

$$\begin{aligned} z_i'(\bar{y}_i - \bar{x}_i\beta - z_i\gamma - \mu_i - \bar{G}_i\nu) &= 0 \\ \vec{G}'(\vec{y} - \vec{X}\beta - \vec{G}\nu) &= 0 \\ \vec{X}'(\vec{y} - \vec{X}\beta - \vec{G}\nu) &= 0 \end{aligned} \quad (5)$$

(where we have used (4) in (5)). To obtain identification (a unique minimizer), additional restrictions must be imposed, as discussed in detail in Abowd *et al.* (2002).

Independent random individual and firm effects Assume now that the vector of the random firm effects, ν , is distributed as

$$\nu \sim \mathcal{N}(0, \sigma_{\nu\nu}I_M),$$

where I_M is the identity matrix of dimension M , and the vector of individual effects, μ , is distributed as

$$\mu \sim \mathcal{N}(0, \sigma_{\mu\mu}I_N).$$

If μ and ν are independent, then

$$\begin{aligned} \bar{y}_i - \bar{x}_i\beta - z_i\gamma - \bar{G}_i\nu &= \bar{\eta}_i + \mu_i \equiv \omega_i, \quad i = 1, \dots, N \\ \vec{y} - \vec{X}\beta - \vec{G}\nu &= \vec{\eta}, \end{aligned} \quad (6)$$

where $\vec{\eta}$ and ω_i are uncorrelated for all i and independent of ν , with $\vec{\eta} \sim \mathcal{N}(0, \sigma_{\eta\eta}I)$,

where I is the identity matrix of dimension $\sum_{i=1}^N (T_i - 1)$, and $\omega_i \sim \mathcal{N}(0, \sigma_{\eta\eta} (T_i^{-1} + \lambda))$, with

$\lambda = \sigma_{\mu\mu}/\sigma_{\eta\eta}$. More compactly, define $\bar{y} = (\bar{y}_1, \dots, \bar{y}_N)'$ and similarly (\bar{x}, \bar{G}) by stacking \bar{x}_i and \bar{G}_i . We can then stack \bar{y} and \bar{y} to obtain

$$\begin{bmatrix} \bar{y} \\ \bar{y} \end{bmatrix} = \begin{bmatrix} \bar{x} \\ \bar{X} \end{bmatrix} \beta + \begin{bmatrix} z \\ 0 \end{bmatrix} \gamma + \begin{bmatrix} \bar{G} \\ \bar{G} \end{bmatrix} \nu + \begin{bmatrix} \bar{\omega} \\ \bar{\eta} \end{bmatrix},$$

where $z = (z'_1, \dots, z'_N)'$ and $\bar{\omega} = (\omega_1, \dots, \omega_N)'$.

Let $\theta = (\beta', \gamma', \sigma_{\eta\eta}, \lambda, \sigma_{\nu\nu})$ denote all parameters to be estimated, and $\theta^{(m)}$ the current estimate of θ (in the m 'th iteration of the estimation algorithm). Furthermore, let $\Omega(\lambda) = \text{diag}(T_1^{-1} + \lambda, \dots, T_N^{-1} + \lambda)$. According to the EM algorithm we can write

$$M(\theta | \theta^{(m)}) = M^{(1)}(\beta, \gamma, \sigma_{\eta\eta}, \lambda | \theta^{(m)}) + M^{(2)}(\sigma_{\nu\nu} | \theta^{(m)}), \quad (7)$$

where

$$\begin{aligned} M^{(1)}(\beta, \gamma, \sigma_{\eta\eta}, \lambda | \theta^{(m)}) &= -\frac{1}{2} \sum_{i=1}^N T_i \ln \sigma_{\eta\eta} - \frac{1}{2} \sum_{i=1}^N \ln \left(\frac{1}{T_i} + \lambda \right) \\ &\quad - \frac{1}{2} \sigma_{\eta\eta}^{-1} \mathbf{E} \left\{ \sum_{i=1}^N \left(\frac{1}{T_i} + \lambda \right)^{-1} (\bar{y}_i - \bar{x}_i \beta - z_i \gamma - \bar{G}_i \nu)^2 | Y; \theta^{(m)} \right\} \\ &\quad - \frac{1}{2} \sigma_{\eta\eta}^{-1} \mathbf{E} \left\{ \left(\bar{y} - \bar{X} \beta - \bar{G} \nu \right)' \left(\bar{y} - \bar{X} \beta - \bar{G} \nu \right) | Y; \theta^{(m)} \right\} \end{aligned} \quad (8)$$

and

$$M^{(2)}(\sigma_{\nu\nu} | \theta^{(m)}) = -\frac{N}{2} \ln |\sigma_{\nu\nu}| - \frac{1}{2} \sigma_{\nu\nu}^{-1} \mathbf{E} \left\{ \nu' \nu | Y; \theta^{(m)} \right\}. \quad (9)$$

In (8)–(9), the expectation is with respect to the latent variables ν conditional on the data Y , and with θ evaluated at $\theta^{(m)}$. Thus $M(\theta | \theta^{(m)})$ is the expected “complete data” log-likelihood, obtained by considering ν as observed random variables and then taking the conditional expectation of this log-likelihood with respect to the latent variables (given Y and the current parameter estimates). It is shown in Dempster *et al.* (1977) that repeated maximization of $M(\theta | \theta^{(m)})$ with respect to θ

generates a sequence $\{\theta^{(m)}\}$, which converges to a stationary point of the likelihood function under very general conditions. Because $M(\theta|\theta^{(m)})$ is quadratic in (μ', ν') , to evaluate the expectations in (8)–(9) we only need to calculate the conditional expectations

$$\widehat{\nu}(\theta^{(m)}) = \mathbf{E} \left\{ \nu \mid Y; \theta^{(m)} \right\}, \quad (10)$$

and the conditional covariance matrix

$$V(\theta^{(m)}) = \mathbf{Var} \left\{ \nu \mid Y; \theta^{(m)} \right\}. \quad (11)$$

We have (see Francke *et al.*, 2010)

$$V(\theta^{(m)}) = \left(\sigma_{\nu\nu}^{-1} I_M + \sigma_m^{-1} \left(\overline{G}' \Omega(\lambda^{(m)})^{-1} \overline{G} + \overline{G}' \overline{G} \right) \right)^{-1}$$

and

$$\widehat{\nu}(\theta^{(m)}) = \sigma_m^{-1} V(\theta^{(m)}) \begin{bmatrix} \overline{G}' \Omega(\lambda^{(m)})^{-1} & \overline{G}' \end{bmatrix} \left(\begin{bmatrix} \overline{y} \\ \overline{y} \end{bmatrix} - \begin{bmatrix} \overline{x} \\ \overline{X} \end{bmatrix} \beta - \begin{bmatrix} z \\ 0 \end{bmatrix} \gamma \right).$$

Because the maximization of $M(\theta|\theta^{(m)})$ is complicated, we suggest modifying the EM algorithm, replacing it with an Expectation Conditional Maximization (ECM) algorithm (see Meng and Rubin, 1993). First, we maximize $M(\theta|\theta^{(m)})$ w.r.t. β and γ given $\lambda = \lambda^{(m)}$. The first-order conditions are given by

$$\begin{aligned} \sum_{i=1}^N z_i' (T_i^{-1} + \lambda^{(m)})^{-1} \left(\overline{y} - \overline{x}_i \beta^{(m+1)} - z_i \gamma^{(m+1)} - \overline{G} \widehat{\nu}(\theta^{(m)}) \right) &= 0 \\ \sum_{i=1}^N \overline{x}_i' (T_i^{-1} + \lambda^{(m)})^{-1} \left(\overline{y} - \overline{x}_i \beta^{(m+1)} - z_i \gamma^{(m+1)} - \overline{G} \widehat{\nu}(\theta^{(m)}) \right) &+ \\ \overline{X}' \left(\overline{y} - \overline{X} \beta^{(m+1)} - \overline{G} \widehat{\nu}(\theta^{(m)}) \right) &= 0. \end{aligned} \quad (12)$$

Then we update $(\sigma_{\eta\eta}, \lambda)$ as follows:

$$(\sigma_{\eta\eta}^{(m+1)}, \lambda^{(m+1)}) = \arg \max_{\sigma_{\eta\eta}, \lambda} r(\sigma_{\eta\eta}, \lambda),$$

where

$$\begin{aligned} r(\sigma_{\eta\eta}, \lambda) &= -\frac{1}{2} \sum_{i=1}^N T_i \ln \sigma_{\eta\eta} - \frac{1}{2} \sum_{i=1}^N \ln \left(\frac{1}{T_i} + \lambda \right) \\ &\quad - \frac{1}{2} \sigma_{\eta\eta}^{-1} \sum_{i=1}^N \left(\frac{1}{T_i} + \lambda \right)^{-1} \left\{ (\bar{y}_i - \bar{x}_i \beta^{(m+1)} - z_i \gamma^{(m+1)} - \bar{G}_i \hat{\nu}(\theta^{(m)}))^2 + \bar{G}_i V(\theta^{(m)}) \bar{G}_i' \right\} \\ &\quad - \frac{1}{2} \sigma_{\eta\eta}^{-1} \left((\bar{\mathbf{y}} - \bar{\mathbf{X}} \beta^{(m+1)} - \bar{\mathbf{G}} \hat{\nu}(\theta^{(m)}))' (\bar{\mathbf{y}} - \bar{\mathbf{X}} \beta^{(m+1)} - \bar{\mathbf{G}} \hat{\nu}(\theta^{(m)})) + \text{tr}(\bar{\mathbf{G}} V(\theta^{(m)}) \bar{\mathbf{G}}') \right). \end{aligned}$$

Finally,

$$\sigma_{\nu\nu}^{(m+1)} = \frac{1}{N} (\hat{\nu}(\theta^{(m)})' \hat{\nu}(\theta^{(m)}) + \text{tr}(V(\theta^{(m)}))). \quad (13)$$

The ECM algorithm then works as follows.

Let $\theta^{(1)}$ be given. For $m = 1, 2, \dots$

(i) The E step: Evaluate $V(\theta^{(m)})$ and $\hat{\nu}(\theta^{(m)})$.

(ii) The CM step: Set

$$\begin{aligned} (\beta^{(m+1)}, \gamma^{(m+1)}) &= \arg \max_{\beta, \gamma} M^{(1)}(\beta, \gamma, \sigma_{\eta\eta}^{(m)}, \lambda^{(m)} | \theta^{(m)}) \\ (\sigma_{\eta\eta}^{(m+1)}, \lambda^{(m+1)}) &= \arg \max_{\sigma_{\eta\eta}, \lambda} M^{(1)}(\beta^{(m+1)}, \gamma^{(m+1)}, \sigma_{\eta\eta}, \lambda | \theta^{(m)}) \\ \sigma_{\nu\nu}^{(m+1)} &= \arg \max_{\sigma_{\nu\nu}} M^{(2)}(\sigma_{\nu\nu} | \theta^{(m)}). \end{aligned}$$

(iii) Set $m = m + 1$, and go to (i) unless $|\theta^{(m+1)} - \theta^{(m)}| < \delta$, for some tolerance level $\delta > 0$ and norm $|\cdot|$. In that case, set $\hat{\theta} = \theta^{(m+1)}$.

Convergence of the above ECM algorithm to a stationary point on the likelihood function follows from Theorem 3 in Meng and Rubin (1993). It follows from the above relations that the estimator with fixed firm effects is a limiting case of the random effects estimator when $\sigma_{\nu\nu}^{-1}$ equals zero, which can be interpreted as assuming a “diffuse” prior for the random effects. See Francke *et al.* (2010) for more details about the relation between the fixed and random effects estimators.

Fixed firm effects and random individual effects Assume now that the individual effects are random and the firm effects are fixed. Then ν is a fixed parameter vector in (8), and there is no conditional expectation involved. Instead, ν must be “maximized out” of (8). The only necessary modification of the conditional maximization algorithm is that in the expression for $r(\sigma_{\eta\eta}, \lambda)$, $V(\theta^{(m)}) = 0$ while $\widehat{\nu}(\theta^{(m)})$, is replaced by $\nu^{(m+1)}$. Moreover, the first-order condition (12) is replaced by the following first-order conditions with respect to $(\beta^{(m+1)}, \gamma^{(m+1)}, \nu^{(m+1)})$:

$$\begin{aligned}
& \sum_{i=1}^N z'_i (T_i^{-1} + \lambda^{(m)})^{-1} \left(\bar{y} - \bar{x}_i \beta^{(m+1)} - z_i \gamma^{(m+1)} - \bar{G} \nu^{(m+1)} \right) = 0 \\
& \sum_{i=1}^N \bar{x}'_i \left(\frac{1}{T_i} + \lambda^{(m)} \right)^{-1} \left(\bar{y} - \bar{x}_i \beta^{(m+1)} - z_i \gamma^{(m+1)} - \bar{G}_i \nu^{(m+1)} \right) + \\
& \bar{X}' \left(\bar{y} - \bar{X} \beta^{(m+1)} - \bar{G} \nu^{(m+1)} \right) = 0 \\
& \sum_{i=1}^N \bar{G}'_i \left(\frac{1}{T_i} + \lambda^{(m)} \right)^{-1} \left(\bar{y} - \bar{x}_i \beta^{(m+1)} - \bar{G}_i \nu^{(m+1)} \right) + \bar{G}' \left(\bar{y} - \bar{X} \beta^{(m+1)} - \bar{G} \nu^{(m+1)} \right) = 0.
\end{aligned} \tag{14}$$

The conditional maximization algorithm then alternates between minimizing $r(\sigma_{\eta\eta}, \lambda)$ and solving (14).

Dependent random individual and firm effects In this case we need to integrate out μ conditional on ν . Thus we must specify the conditional distribution

$$\mu | \nu \sim \mathcal{N}(A(\rho)\nu, \Sigma(\rho)),$$

where ρ is a vector of free parameters. In the general case, where

$$\begin{bmatrix} \mu \\ \nu \end{bmatrix} \sim \mathcal{N} \left(0, \begin{bmatrix} \Sigma_{aa} & \Sigma_{ab} \\ \Sigma'_{ab} & \Sigma_{bb} \end{bmatrix} \right),$$

we have

$$\begin{aligned}
E(\mu | \nu) &= \Sigma_{ab} \Sigma_{bb}^{-1} \nu = A(\rho) \nu \\
Var(\mu | \nu) &= \Sigma_{aa} - \Sigma_{ab} \Sigma_{bb}^{-1} \Sigma'_{ab} = \Sigma(\rho).
\end{aligned}$$

To obtain a feasible model, some simplifications must be made. Let $\nu_{i(t^1)}, \nu_{i(t^2)}, \dots, \nu_{i(t^{m_i})}$ denote the m_i distinct elements of $\nu_{i(1)}, \dots, \nu_{i(T_i)}$. Henceforth we assume that $(\mu_i, \nu_{i(t^1)}, \nu_{i(t^2)}, \dots, \nu_{i(t^{m_i})})$ have a joint normal distribution:

$$\begin{bmatrix} \mu_i/(\sigma_{\mu\mu})^{1/2} \\ \nu_{i(t^1)}/(\sigma_{\nu\nu})^{1/2} \\ \vdots \\ \nu_{i(t^{m_i})}/(\sigma_{\nu\nu})^{1/2} \end{bmatrix} \sim \mathcal{N} \left(\begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \tilde{\rho} & \cdots & \tilde{\rho} \\ & 1 & 0 & \cdots \\ & & \ddots & 0 \\ & & & 1 \end{bmatrix} \right). \quad (15)$$

Then it follows that $E(\mu_i|\nu) = \left(\frac{\sigma_{\mu\mu}}{\sigma_{\nu\nu}}\right)^{\frac{1}{2}} \tilde{\rho} \sum_{j=1}^{m_i} \nu_{i(t^j)} \equiv \rho T_i G_i \nu$, with $\rho = \left(\frac{\sigma_{\mu\mu}}{\sigma_{\nu\nu}}\right)^{\frac{1}{2}} \tilde{\rho}$, and $Var(\mu_i|\nu) = \sigma_{\mu\mu}(1 - m_i \tilde{\rho}^2)$. We henceforth ignore terms of order $O(\tilde{\rho}^2)$, assuming $m_i \tilde{\rho}^2 \approx 0$. This assumption conforms with most estimates of $\tilde{\rho}$ in the literature based on fixed effects estimators; see Andrews *et al.* (2008), who find that $|\tilde{\rho}| \approx 0.05$. The only modification needed then is that the (row) vector \bar{G}_i is replaced by $(1 + \rho T_i) \bar{G}_i$. Conditional maximization with respect to ρ must be performed by augmenting θ by ρ and extending the ECM algorithm by a separate maximization of $M^{(1)}(\beta, \gamma, \sigma_{\eta\eta}, \lambda, \rho)$ with respect to ρ .

3 Application: Wage equation estimation

We consider the following wage equation:

$$\log(W_{ijt}) = Z_i \gamma + X_{it} \beta + \kappa_t + \mu_i + \nu_j + \eta_{ijt}, \quad (16)$$

where W_{ijt} is the annual wage for (full-time employee) i employed in firm j in year t , and the variables in the two vectors of explanatory variables are

$$Z_i = (\text{years of schooling, type of education-dummies, gender})$$

$$X_{it} = (\text{powers of experience up to the third order, labor market area dummies}).$$

In the notation of the previous section, we have $y_{ijt} = \log(W_{ijt})$, $z_i = Z_i$, $x_{it} = (X_{it}, 1(t=1), \dots, 1(t=T))$, where $1(t=s)$ is one if $t=s$, and zero otherwise. The symbol κ_t represents fixed time effects.

The specification in (16) is rather general and may be specialized in various ways. We consider three main specifications of the wage equation. For all three, the unobserved individual-specific effects are treated as random effects, while the unobserved firm-specific effects are either ignored or formulated as random effects or as fixed effects. These three specifications are denoted RENO, RERE and REFE, respectively.⁷ Finally, we compare our estimates to a specification where we treat both the individual- and the firm-specific effects as fixed, denoted FEFE. The disadvantage is then, of course, that the γ parameters are not identified, including the coefficients of the education variables. In a wage model specified on matched employer–employee data, with the main focus on returns to education, it is necessary to model unobserved individual heterogeneity as a random effect. Furthermore, one might argue that when individuals move from one firm to another, independent of reason, individual choice or plant closures, such a stochastic process is not well described by individual fixed effects. Thus, the FEFE regression should only be seen as a robustness check.⁸

The initial sample included 241,904 observations, for 53,665 individuals. The sample covered the period 1995–2006 and was collected for individuals and firms in the Norwegian machinery industry (NACE 29). In total, there were 2,593 firms in the initial sample. For those individuals whose length of education changed over the

⁷However, (16) may be said to be somewhat asymmetric in that whereas we allow for the influences of individual-specific observed variables, we do not add firm-specific observed variables. In the empirical part of the paper, we conduct a robustness check where we include mean employment of the firm as an additional regressor. Some contributions to the literature that estimate wage equations on employer–employee data have allowed for firm effects; for example, see the analyses by Lallemand *et al.* (2005), Plasman *et al.* (2007) and Heyman (2007).

⁸This is the specification considered in the seminal paper by Abowd *et al.* (1999). However, these authors seem to disregard unidimensional variables from the outset in their analysis.

sample period, we retained only the observations with maximum length of education. We included only individuals whose annual earnings were between 50,000 and 3,500,000 NOK (fixed prices).⁹ Labor market experience is represented by potential experience, that is, age minus years of schooling minus seven years. The definition of the labor market region dummies is based on characteristics such as size and centrality.¹⁰ Mainly workers with the following three types of education are represented in the chosen industries: education in “General Programs”, “Business and Administration” and “Natural Sciences, Vocational and Technical subjects”. Only these categories are therefore represented by education-type dummies in the model. The earnings measure used was total annual taxable (full-time) labor income. Because the earnings measure reflects annual earnings, observations where employment relationships began or terminated within the actual year were excluded. Holders of multiple jobs and individuals who received unemployment benefits or participated in active labor market programs were excluded. It was also required that each individual have two or more observations after the abovementioned exclusion criteria were applied. After the data were cleaned as described above, the sample included 201,833 observations, 36,183 individuals and 2,178 firms over the period 1995–2006.

Because we focus on models with both individual- and firm-specific unobserved effects, it is important that a substantial proportion of the individuals are observed in at least two different firms over the period they occur in the sample. This is necessary to identify the unobserved firm effects, regardless of whether these are specified as random or fixed. In fact, in our main data sample we only include individuals that are movers; that is, they change employer at least once during the period they are in the sample. There are 9,400 individuals, with a total of 70,509 observations, who move from one employer to another at least once, and thus help

⁹1 Euro \approx 8 NOK.

¹⁰See http://www.ssb.no/english/subjects/06/sos.110_en.pdf.

to identify the unobserved firm effects. Finally, from this data sample we randomly draw observations for 2,021 individuals with a total of 15,415 observations and 4,476 unique worker–firm combinations. This random draw was conducted to accelerate estimation. Tables A1 and A2 provide some information about the unbalanced panel data set. The individuals are observed from a minimum of two to a maximum of 12 years. On average there are 7.6 observations per individual.

Empirical results Table 1 contains estimation results of the wage equation under different assumptions with respect to the treatment of unobserved individual- and firm-specific heterogeneity.

[Table 1 here]

In the first specification, column 1, an individual random effects model is used and no firm controls are included; that is, these are the results for the RENO model. The estimated return to an additional year of education is 0.073. This estimate seems somewhat high. Turning to the models with firm effects outlined in Section 3, we find that the returns to education become clearly smaller for the REFE specification (0.063), and less so for the RERE model (0.070). These findings may indicate that models with random firm effects are misspecified, being contaminated by omitted variable bias. From Table 1 we notice that the parameter estimates obtained for RENO and RERE are fairly equal. As mentioned in the introduction, this is to be expected because they only differ in the parametrization of the covariance matrix of the gross error terms. Furthermore, if we consider the RENO and the REFE estimates together, we find that the latter is one percentage point smaller (13 percent) than the former, a difference that is quite substantial. An estimated difference of 0.01 is also relatively large when we take statistical uncertainty into account (the standard error is 0.002).

The parameter estimates for the experience coefficients do not vary greatly between the three models. The maximum return to experience is found to be after 25–30 years of experience, and is more or less flat thereafter, as Figure A1 in the appendix shows. The local minimum at about 45 years of experience is not a substantial feature, reflecting the fact that only a few workers (less than 2.5 percent) have such long experience. The estimates of the male dummy are greater than 0.20, showing that the gender wage gap is significant. This is quite large and should be investigated further.¹¹ None of the education-type parameters is found to be individually statistically significant for the RENO model. For the last two models in which we control for unobserved firm effects, RERE and REFE, we find the dummies for “General Programs” and “Business and Administration” to be statistically significant. These two groups include managers and administrative personnel.

The estimate of the correlation coefficient between the unobserved individual and firm effects in the RERE model is estimated as 0.13 and is highly significant (standard error < 0.01).¹² This estimate is close to the 0.11 reported by Abowd *et al.* (1999), who used a FEFE specification. Torres *et al.* (2010) report a somewhat higher positive estimate. On the other hand, Andrews *et al.* (2008), Grütter and Lalive (2009) and Cornelißen and Hübler (2011) report negative estimates. However, the interpretation and comparison of these results in view of the substantial question of sorting are not straightforward both because of the theoretical considerations outlined by Eeckhout and Kircher (2010) and because our estimate relies on a specification whereby both unobserved individual and firm heterogeneity are represented by random effects, whereas the above studies apply specifications in which both components are assumed to be fixed effects. The fixed effects capture

¹¹In this paper, we do not focus on gender differences when modeling wages. We only consider the male dummy as a control variable.

¹²We also estimated a specification of the RERE model in which we forced the correlation between the firm-specific and the individual-specific terms to be zero. This restriction leads to only very small changes in the coefficient estimates reported in Table 1, column (2).

the influence of all one-dimensional observed variables, whereas the random effects specification only captures heterogeneity beyond what is already accounted for by the inclusion of the time-invariant regressors.

Although our approach also covers the standard model with fixed individual-specific effects, FENO, and the FEFE model, these are of minor interest given that we are interested in estimating the return to education, which is not identified in the presence of fixed individual effects.¹³ Moreover, it is not possible to identify the return to experience when we allow a general time trend (time dummies) because (potential) experience increases linearly over time and therefore becomes collinear with the time dummies and the dummies representing the fixed individual effects.¹⁴¹⁵

We tested the RERE model against the REFE model (i.e., fixed firm effects) using a Hausman test, in which the null hypothesis is that the RERE model is correct. The test statistic exceeded 95 (with 25 degrees of freedom), and the p-value was practically equal to zero. Because Hausman tests routinely reject the random effect specification in large samples, this test may not be very informative in our case. The large estimated value of $\sigma_{\nu\nu}/\sigma_{\eta\eta}$ compared with $\sigma_{\mu\mu}/\sigma_{\eta\eta}$, reported in Table 1 (i.e., 16.1 vs. 1.07), shows that firm effects have a more dispersed distribution than do individual effects. Note that in the limiting case when $\sigma_{\nu\nu}/\sigma_{\eta\eta}$ tends to infinity, we obtain the REFE model. As mentioned above, neither the parameters

¹³For alternative algorithms of estimating the FEFE model, see Cornelissen (2008) and Guimarães and Portugal (2010).

¹⁴We also ran the RENO model in which the stayers are added to the sample. In this way, we found the estimate of the slope parameter of education length to be smaller, 0.0705, and the estimate of the parameter attached to the first-order power of experience to be 0.0415, that is, somewhat larger than the estimates reported in Table 1.

¹⁵We performed a robustness check where we include mean number of employees per firm in the RENO model. This is in the spirit of Mundlak (1978). The estimation results obtained using this formulation are very similar to the results for the RENO and RERE specifications reported in Table 1. The estimate of the return to education is 0.0708, compared with 0.0729 in the RENO model and 0.0703 in the RERE model. One may think of the Mundlak approach as an alternative to including fixed firm effects, but because the estimate of the return to schooling deviates from the estimate obtained using the REFE model, more time-invariant firm-specific variables may be needed to obtain better conformity.

corresponding to length of education nor those corresponding to experience are identified in the model with fixed individual effects; hence we cannot test REFE versus FEFE (because they do not contain the same explanatory variables). In fact, the time invariance of the education variable makes the use of individual-specific fixed effects models inappropriate. One alternative would have been to estimate the wage equation using the estimator put forward by Hausman and Taylor (1981). However, to use this approach it is necessary to identify which variables are correlated and which are uncorrelated with the unobserved individual-specific effect. In our case it would be rather speculative to make such distinctions.

4 Concluding remarks

In this paper, we considered a general regression model with an unobserved random effect corresponding to the main observation unit and an unobserved effect corresponding to another type of unit with which the main observation unit is matched at a given point in time. In an application, we examined different specifications of real wage equations requiring access to employer–employee panel data. Such data enable controls for both unobserved individual- and firm-specific effects. Earlier contributions in this area—of those Abowd *et al.* (1999) being the best known and most often cited—have stuck to a specification with both fixed individual- and firm-specific effects. However, a feature of such a specification is that one cannot identify effects on the real wage of one-dimensional variables, such as length and type of education or gender. Thus, it may be worthwhile to consider random effects specifications, as we have done in this paper. To estimate the model we applied the Helmert transformation on the wage equation to sweep out the individual-specific random effects. To obtain estimates of the unknown parameters, the profile likelihood was maximized using the Expectation Conditional Maximization algorithm.

Using this approach, we find the estimate of return to education to become more than 10 percent smaller when, in addition to controlling for unobserved individual specific effects, we control for fixed firm-specific effects.

References

- [1] Abowd, J.M., Kramarz, F., and Margolis, D.N. (1999), “High Wage Workers and High Wage Firms,” *Econometrica*, 67, 251–333.
- [2] Abowd, J.M., Creecy, R.H., and Kramarz, F. (2002), “Computing Person and Firm Effects Using Linked Longitudinal Employer–Employee Data (2002),” Technical Paper 2002–2006. US Census Bureau.
- [3] Andrews, M.J., Gill. L., Schank, T., and Upward, R. (2008), “High Wage Workers and Low Wage Firms,” *Journal of the Royal Statistical Society, Series A*, 171, 673–697.
- [4] Arellano, M., and Bover, O. (1995), “Another Look at the Instrumental Variables Estimation of Error Component Models,” *Journal of Econometrics*, 68, 29–51.
- [5] Balestra, P., and Krishnakumar, J. (2008), “Fixed Effects Models and Fixed Coefficient Models,” in *The Econometrics of Panel Data: Fundamentals and Recent Developments in Theory and Practice*, eds. Mátyás, L. and P. Sevestre, Berlin: Springer, pp. 23–48.
- [6] Blanchard, P., Gaigné, C., and Mathieu, C. (2008), “Foreign Direct Investments: Lessons from Panel Data,” in *The Econometrics of Panel Data: Fundamentals and Recent Developments in Theory and Practice*, eds. Mátyás, L. and P. Sevestre, Berlin: Springer, pp. 663–696.
- [7] Cornelissen, T. (2008), “The Stata Command felsdvvreg to Fit a Linear Model with Two High-Dimensional Fixed Effects,” *Stata Journal*, 8, 170–189.
- [8] Cornelißen, T., and Hübler, O. (2011), “Unobserved Individual and Firm Het-

erogeneity in Wage and Job Duration Functions: Evidence from German Linked Employer–Employee Data,” *German Economic Review* (forthcoming).

- [9] Demidenko, E. (2004), *Mixed Models: Theory and Applications*, Hoboken, New Jersey: Wiley.
- [10] Dempster, A.P., Laird, N.M., and Rubin, D.B. (1977), “Maximum Likelihood from Incomplete Data via the EM Algorithm (with discussion),” *Journal of the Royal Statistical Society, Series B*, 39, 1–38.
- [11] Eeckhout, J., and Kircher, P. (2010), “Identifying Sorting – In Theory,” LSE Research Online.
- [12] Francke, M.K., Koopman, S.J., and De Vos, A.F. (2010), “Likelihood Functions for State Space Models with Diffuse Initial Conditions,” *Journal of Time Series Analysis*, 31, 407–414.
- [13] Godager, G., and Biørn, E. (2010), “Does Quality Influence Choice of General Practitioner? An Analysis of Matched Doctor–Patient Panel Data,” *Economic Modelling*, 27, 842–853.
- [14] Goux, M., and Maurin, E. (1999), “Persistence of Interindustry Wage Differentials Using Matched Worker-Firm Data,” *Journal of Labor Economics*, 17, 492–533.
- [15] Grütter, M., and Lalive, R. (2009), “The Importance of Firms in Wage Determination,” *Labour Economics*, 16, 149–160.
- [16] Guimarães, P., and Portugal, P. (2010), “A Simple Feasible Procedure to Fit Models with High-Dimensional Fixed Effects,” *Stata Journal*, 10, 628–649.

- [17] Hausman, J.A., and Taylor, W.E. (1981), “Panel Data and Unobservable Individual Effects,” *Econometrica*, 49, 1377–1398.
- [18] Heyman, F. (2007), “Firm Size or Firm Age? The Effects on Wages Using Matched Employer–Employee Data,” *Labour*, 21, 237–263.
- [19] Ioannidou, V., and Ongena, S. (2010), “‘Time for a Change’: Loan Conditions and Bank Behavior when Firms Switch Banks,” *Journal of Finance*, 65, 1847–1877.
- [20] Keane, M.P., and Runkle, D.E. (1992), “On the Estimation of Panel-Data Models when Instruments are not Strictly Exogenous,” *Journal of Business and Economic Statistics*, 10, 1–9.
- [21] Lallemand, T., Plasman, R., and Rycx, F. (2005), “Why Do Large Firms Pay Higher Wages? Evidence from Matched Worker–Firm Data,” *International Journal of Manpower*, 26, 705–723.
- [22] Meng, X.-L., and Rubin, D.B. (1993), “Maximum Likelihood Estimation via the ECM Algorithm: A General Framework,” *Biometrika*, 80, 267–278.
- [23] Mundlak, Y. (1978), “On the Pooling of Time Series and Cross Section Data,” *Econometrica*, 46, 69–85.
- [24] Plasman, R., Rycx, F., and Tojerow, I. (2007), “Wage Differentials in Belgium: The Role of Worker and Employer Characteristics,” *Cahiers Economiques de Bruxelles*, 50, 11–40.
- [25] Searle, S.R., Casella, G., and McCulloch, C.E. (1992), *Variance Components*, New York: Wiley.

- [26] Torres, R., Portugal, P., Addison, J.T., and Guimarães, P. (2010), “The Sources of Wage Variation: An Analysis Using Matched Employer Employee Data,” Working Papers 25/2010, Banco de Portugal.

Tables and figures

Table 1: Empirical results

Individual effects	RE	RE	RE
Firm effects	NO	RE	FE
eduleng	0.0729 (0.0026)	0.0703 (0.0023)	0.0632 (0.0022)
exper	0.0355 (0.0025)	0.0359 (0.0023)	0.0375 (0.0024)
exper ² /100	-0.1018 (0.0116)	-0.1048 (0.0107)	-0.1121 (0.0108)
exper ³ /10000	0.0922 (0.0158)	0.0969 (0.0146)	0.1059 (0.0146)
male	0.2148 (0.0176)	0.2167 (0.0151)	0.2327 (0.0146)
Education type			
General Programs	0.0445 (0.0261)	0.0818 (0.0226)	0.05 (0.0222)
Business and Administration	0.0522 (0.0297)	0.0884 (0.0243)	0.0546 (0.0238)
Sci&Tech	-0.0118 (0.0247)	0.0201 (0.0205)	-0.0075 (0.0202)
$\sigma_{\mu\mu}/\sigma_{\eta\eta}$		1.07	1.36
$\sigma_{\nu\nu}/\sigma_{\eta\eta}$	–	16.1	∞
Corr(Ind. effect, Firm effect)		0.13 (0.008)	
No. of obs	15,415	15,415	15,415
No. of individuals	2,021	2,021	2,021
No. of firms	770	770	770
Unique worker–firm comb.	4,476	4,476	4,476

Notes: Standard errors in parentheses. Year and labor market area effects are accounted for in all the models. RE, FE and NO denote random, fixed and *no* (individual or firm) effects, respectively.

Appendix

Table A1: The unbalancedness of the panel data

Number of years a person is in the sample	Number of persons in the sample	Number of observations
2	87	174
3	153	459
4	203	812
5	184	920
6	168	1,008
7	165	1,155
8	173	1,384
9	180	1,620
10	171	1,710
11	271	2,981
12	266	3,192
Sum	2,021	15,415

Table A2. Overview of number of firms in workers' employment history

Number of firms	Number of individuals having worked in the indicated number of firms
2	1,650
3	312
4	55
5	4
Total	2,021

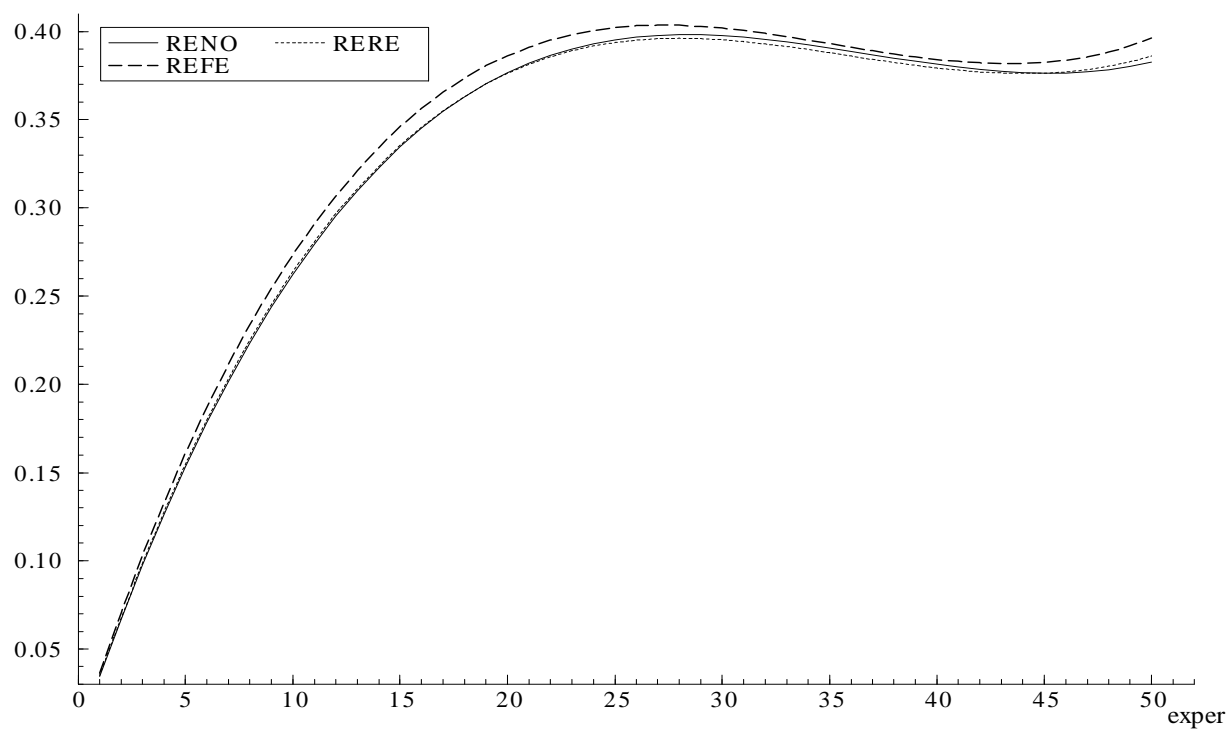


Figure A.1: The partial effect of experience on expected log-wage. Estimates for three model specifications.