

A Simple Method to Estimate Large Fixed Effects Models Applied to Wage Determinants

Nikolas Mittag, CERGE-EI

September 2019

Abstract

Models with high-dimensional sets of fixed effects are frequently used to examine, among others, linked employer-employee data, student outcomes and migration. Estimating these models is computationally difficult because of the high-dimensional design matrix. I present a simple algorithm to compute the OLS estimates of large two-way fixed effects (TWFE) and match effect models including estimates of the fixed effects. The algorithm simplifies specification tests and variance estimation even with multi-way clustered errors. An application using German linked employer-employee data illustrates key advantages of the algorithm: Omitting match effects substantially affects estimates including the gender wage gap. Analyzing the estimated fixed effects suggest that firm fixed effects are the main channel through which job transitions drive wage dynamics, which underlines the importance of firm heterogeneity for labor market dynamics.

JEL Classification: J31, J63, C23, C63

Keywords: multi-way fixed effects, match effects, linked employer-employee data, wage dynamics

I would like to thank Dan Black, Randy Filer, Jeffrey Grogger, Stepan Jurajda, Francis Kramarz, Bruce Meyer and participants at the IAB, RWI, EALE and the Czech Economic Society for comments and suggestions as well as Stefan Bender and the staff at the IAB for the excellent cooperation with the data. Marin Drlje provided great research assistance. All errors are mine. Address: CERGE-EI, joint workplace of Charles University and the Economics Institute of the Czech Academy of Sciences, Politických vězňů 7, 110 00 Prague, Czech Republic. Email: nikolasmittag@posteo.de

1. Introduction

Following the seminal article on the two way fixed effects model (TWFE) by Abowd, Kramarz and Margolis (1999), models with large sets of fixed effects are frequently used to examine a wide range of topics in labor economics (see Abowd, Kramarz and Woodcock 2008 for an overview), education (e.g. Jacob and Lefgren 2008, Kramarz, Machin and Ouazad 2008, Jackson 2013), health (e.g. Bennett, Hung and Lauderdale 2015) and migration (Grogger and Hanson 2011). The match effects model extends the TWFE by including the interaction of the two fixed effects, e.g. a match or job fixed effect. Match effects are of interest whenever the way in which agents are paired with each other matters, which is likely in many interactions of heterogeneous populations. For example, theories of matching between employers and firms (e.g. Jovanovic 1979, Mortensen 1978) usually allow for complementarities between workers and firms, which could be captured by adding a match effect to the firm and individual effects of the TWFE. However, the computational complexity of both models makes specification and estimation difficult. Even in the simpler TWFE the number of fixed effects often exceeds one million, so that regression coefficients cannot be computed from the standard formula, because it is impossible to invert the cross-product of the design matrix.

This paper introduces a fast and simple method to estimate the TWFE and the match effects model making use of the fact that after purging all covariates, formulas for the inverse of partitioned matrices yield a computationally simple solution for one set of fixed effects. This simplification makes the remaining parameters easy to compute. The algorithm I propose in this paper yields the exact OLS solutions including estimates of the fixed effects. It computes the variance matrix of the estimated slope coefficients (but SEs of the fixed effects have to be bootstrapped if needed) even if errors are correlated across observations. A key advantage of the algorithm is that its computational complexity grows in the number of firms (i.e. the dimension of the smaller set of fixed effects) rather than in the number of individuals and firms.

Thereby, the algorithm I propose solves or greatly simplifies the computational problem. I analyze wage determinants using linked employer-employee data from Germany to illustrate key advantages of the algorithm. First, the algorithm facilitates specification tests in two ways. It reduces the computational burden of estimating both the TWFE and the match effects, which simplifies tests for the presence of match effects. The omission of match effects can bias estimates if the pairing of agents matter, which is common in applications (e.g. Garen 1989, Jackson 2013, Woodcock 2008, 2015). The variance matrix of the estimated slopes computed by the algorithm is crucial for significance and model specification tests, but often requires bootstrapping or partial regression. Both strategies require solving the estimation problem repeatedly, which amplifies the computational problem. The application to German data shows that excluding match effects affects estimates of key parameters, as among others the unexplained part of the gender wage gap doubles. Estimating the variance matrix of the slopes makes it possible to test for changes in these slope parameters. Using a Hausman-type test shows that match effects matter for the slope parameters, even though they only explain a relatively small share of the variation in wages (Card, Heining and Kline 2013).

In addition, the method introduced here simplifies the computation of firm-, individual- and (if relevant) match components. Estimates of the fixed effects are often of substantive interest themselves. Several parameters of interest can be calculated from the estimates of the individual and firm fixed effects (e.g. Abowd, Kramarz, Roux 2006) and teacher fixed effects are commonly used as measures of value added (e.g. Rockoff 2004). Estimates of the match effects are of interest in search models (Woodcock 2015), matching of teachers and schools (Jackson 2013) as well as international migration (Grogger and Hanson 2011). Examining what predicts the firm and match fixed effects after job transitions in the German data underlines that firms are not only of importance for wage dispersion (Card, Heining and Kline 2013), but also for wage dynamics.

The next section describes the two models, estimation problems and common methods. Part 3 introduces the algorithm. Part 4 illustrates key advantages of the algorithm by analyzing wage determinants.

2. The Two-Way Fixed Effects and the Match Effects Model

A common specification in panel data models is the two-way fixed effects model which includes a set of fixed effects for primary units indexed by $i = 1, \dots, N$ and secondary units indexed by $j = 1, \dots, J$. I define the secondary unit as the smaller unit in the sense that $J < N$ without loss of generality. For simplicity, I discuss these models in terms of a regression of (log) wages on covariates that includes fixed effects for individuals, firms, and possibly individual-firm matches. Let T_i indicate the number of observations on individual i . Similarly, let F_j stand for the number of observations on firm j , so that the total number of observations is $N^* = \sum_{i=1}^N T_i = \sum_{j=1}^J F_j$. The model allows both T_i and F_j to vary between units, so the panel does not have to be balanced. The TWFE model can be expressed as

$$\underset{N \times 1}{y} = \underset{N \times K}{X} \underset{K \times 1}{\beta} + \underset{N \times N}{D_\theta} \underset{N \times 1}{\theta} + \underset{N \times J}{D_\psi} \underset{J \times 1}{\Psi} + \underset{N \times 1}{\varepsilon} \quad (1)$$

Where y is a vector of outcomes of individual i at firm j at time t , y_{ijt} . X is a matrix of observable time-varying covariates, D_θ is the matrix of indicators for individuals, D_ψ is the matrix of indicators for firms and ε is the vector of unobservables ε_{ijt} . The parameters of the model are β , the vector of slopes, θ , the vector of individual fixed effects θ_i and Ψ , the vector of firm fixed effects ψ_j . The match effects model (see e.g. Woodcock 2008, 2015) extends the TWFE by adding the interaction between the two fixed effects:

$$\underset{N \times 1}{y} = \underset{N \times K}{X} \underset{K \times 1}{\beta} + \underset{N \times N}{D_\theta} \underset{N \times 1}{\theta} + \underset{N \times J}{D_\psi} \underset{J \times 1}{\Psi} + \underset{N \times S}{D_\lambda} \underset{S \times 1}{\lambda} + \underset{N \times 1}{\varepsilon} \quad (2)$$

D_λ is the matrix of indicators for matches between the two units and λ is the vector of match fixed effects λ_s , where $s = 1, \dots, S$ is for convenience only as it is determined by i and j . To ease notation, define

$$\begin{aligned} T &= D'_\theta D_\theta = \text{diag}(T_1, \dots, T_N) \\ F &= D'_\psi D_\psi = \text{diag}(J_1, \dots, J_J) \\ M &= D'_\theta D_\psi \end{aligned} \quad (3)$$

So that T is an $N \times N$ diagonal matrix with the number of observations on individual i as the i^{th} diagonal element and F is a $J \times J$ diagonal matrix with the number of observations on firm j as the j^{th} diagonal element. Element (i, j) of the $N \times J$ matrix M indicates how many periods individual i worked for firm j . Abowd, Creecy, and Kramarz (2002) discuss identification in the TWFE. They show that all worker and firm effects within each connected group (groups of firms with realized mobility) are identified up to one normalization: The level of one set of fixed effects is only identified relative to the other set of fixed effects in each group. The match effects model additionally includes the interactions between firm and individual fixed effects. This match effect is nested both within the firm and within the individual fixed effect. Woodcock (2015) provides a thorough discussion of identification. Most importantly, the mean match effect for each individual and firm is not identified. Intuitively, average match quality is an invariant characteristic of a firm or individual by construction, so it cannot be separated from the person and firm effect. Consequently, we can only learn from the data how a specific match differs from other matches of the same firm and from other matches of the individual.¹ Therefore, match effects are usually constrained to sum to zero for each individual and each firm. This normalization has convenient computational properties and other normalizations are easily to implement. It makes the match effects exactly orthogonal to θ and Ψ by construction, but they are not necessarily orthogonal to X . Thus, omitting match effects in a model without time-varying regressors yields numerically identical estimates of θ and Ψ , but if the model includes X omitting them is likely to affect $\hat{\beta}$ as well as $\hat{\theta}$ and $\hat{\Psi}$.

Estimating the models above by inverting the cross-product of the data matrix is computationally infeasible, because it requires the inverse of a $(K + N + J) \times (K + N + J)$ matrix for the TWFE ($(K + N + J + S) \times (K + N + J + S)$ for the match effects model). In typical datasets, this number easily

¹ Many questions concern comparisons within firm or individual, so the estimated match effects are informative. The data and hence the estimated match effects contain no information on other questions, such as correlations between match quality and time-invariant attributes or comparing match quality across firms or individuals.

exceeds one million. Calculations with matrices of this size require infeasible amounts of memory and time.² For an overview of computational approaches and alternative estimators (such as random or mixed effects), see Abowd, Kramarz and Woodcock (2008) and Andrews, Schank and Upward (2006). Woodcock (2015) shows that the match effect model can be reduced to solving a problem of size $N + J$ by first calculating match means and then decomposing them into the three sets of fixed effects. So both models can be estimated by solving a problem of similar size, but this problem is too large to compute estimates from the standard OLS formulas, because it is infeasible to invert the cross-product of the data matrix.

One approach to reduce computational complexity is to iteratively solve the normal equations, without inverting the cross-product matrix. For example, the conjugate gradient algorithm (CGA, see Abowd, Creedy and Kramarz 2002, Ouazad 2007) and alternating projections (Guimarães and Portugal 2010, Gaure 2013, Correia 2014) iteratively search for the OLS estimates of the slopes and fixed effects. Iterative solutions require applying the algorithm repeatedly to estimate standard errors by partial regression or the bootstrap, which exacerbates the problem that iterative methods are very slow. Another approach reduces the dimensionality of the matrix to be inverted by performing a first-difference (e.g. Abowd, Kramarz and Margolis 1999) or within-transformation (e.g. Andrews, Schank and Upward 2006, Cornelissen 2008) on the larger set of fixed effects. This transformation reduces the size of the matrix to $K + J$, but both approaches still have to invert a very large matrix.

3. A Computationally Simple Estimation Strategy

The algorithm I present below uses a transformation to obtain the slopes, reduces the dimensionality of the matrix and then solves a system of linear equations rather than inverting a matrix. The algorithm proceeds in three steps: First, compute the slope parameters $\hat{\beta}$. Second, compute the smaller set of fixed

² E.g., a matrix with one million rows and columns stored in double precisions requires 8 TB of working memory. Matrix inversion requires computation of cubic order. Computation would still be infeasible if the current lower bound in Le Gall (2014) were practicable. Many algorithms only need to compute matrix factorizations (such as the Cholesky-factorization), which slightly alleviates the computational problem, but mainly improves numerical stability. For simplicity, I will refer to the need to “invert” the matrix throughout the paper.

effects. Third, compute the remaining fixed effects. I first discuss the estimation of the match effects model, then show how to apply it to the TWFE model. I then discuss variance estimation and briefly compare the algorithm to other strategies.

The first step for the match effects model is to compute $\hat{\beta}^{ME}$, the slopes defined by eq. 2 which can be done by an OLS regression of deviations of y and x from their match means. This procedure is equivalent to running a partial regression. Therefore, the estimates, residuals and standard errors (corrected for the degrees of freedom) are numerically identical to those that would be obtained from OLS estimation of the full model (Yule 1907). To see how calculating $\hat{\beta}^{ME}$ simplifies computing the smaller set of fixed effects in the second step, note that $\hat{\beta}^{ME}$ is the OLS estimate, so the OLS estimates of the fixed effects can be obtained by a regression of $\tilde{y} = y - X\hat{\beta}^{ME}$ on the individual and firm fixed effects.³ The normalization that match effects sum to zero within each firm and individual allows one to omit the match effects in this regression because the model does not contain X .⁴ Note that this simplification requires the first step of purging the effect of X , because the match effects need not be exactly orthogonal to the firm and individual effects after conditioning on the covariates. Estimates from a regression of \tilde{y} on the individual and firm fixed effects are then given by the standard formula:

$$\begin{pmatrix} \hat{\theta}^{ME} \\ \hat{\Psi}^{ME} \end{pmatrix} = \begin{pmatrix} T & M \\ M' & F \end{pmatrix}^{-1} \begin{pmatrix} D_{\theta}' \tilde{y} \\ D_{\Psi}' \tilde{y} \end{pmatrix} = \begin{pmatrix} T & M \\ M' & F \end{pmatrix}^{-1} \begin{pmatrix} T \tilde{\bar{y}}_i \\ F \tilde{\bar{y}}_j \end{pmatrix} \quad (4)$$

where $\tilde{\bar{y}}_i$ and $\tilde{\bar{y}}_j$ are vectors of individual and firm means of \tilde{y}_{it} and T, F and M are defined by eq. 3.

The second step is to calculate the firm fixed effects (or, more generally, the smaller set of fixed effects).

To do so, one only needs the lower blocks of the inverse of the partitioned matrix. These blocks can be obtained by applying the formula for the inverse of a partitioned matrix (see Theil 1971 section 1.2):

³ For proof, see e.g. Greene (2008 p. 27) or appendix A.

⁴ Proof: In a regression without time-varying covariates X , the omitting the match effects results in the usual omitted variable bias: $[D_{\theta} \ D_{\Psi}]' [D_{\theta} \ D_{\Psi}]^{-1} [D_{\theta} \ D_{\Psi}]' D_{\lambda} \lambda$. This bias is 0 by construction, because $D_{\theta}' D_{\lambda} \lambda = 0$ and $D_{\Psi}' D_{\lambda} \lambda = 0$ hold exactly within sample by normalizing match effects to sum to zero within individual and firm.

$$\begin{aligned}
\hat{\Psi}^{ME} &= [-(F - M'T^{-1}M)^{-1}M'T^{-1} \quad (F - M'T^{-1}M)^{-1}] \begin{bmatrix} T\bar{\tilde{y}}_i \\ F\bar{\tilde{y}}_j \end{bmatrix} \\
&= (F - M'T^{-1}M)^{-1}F\bar{\tilde{y}}_j - (F - M'T^{-1}M)^{-1}M'\bar{\tilde{y}}_i \\
&= (F - M'T^{-1}M)^{-1}(F\bar{\tilde{y}}_j - M'\bar{\tilde{y}}_i)
\end{aligned} \tag{5}$$

T and F are diagonal matrices and M has at most S non-zero elements, so they are sparse. All three matrices contain only integers. Consequently, these expressions are simple to evaluate. The only remaining computational difficulty is inverting the matrix in the first brackets. It is of size $J \times J$ (and remains symmetric positive definite), so it is already much smaller than the $(K + N + J + S) \times (K + N + J + S)$ matrix in the standard OLS formula. Inverting it directly yields estimates of the firm fixed effects and their standard errors. One can also use iterative algorithms such as the CGA to solve the much smaller $J \times J$ system of equations implied by equation 5.

The third step of calculating the remaining fixed effects is simple, because residuals sum to zero for each individual, firm and match. Therefore, the individual fixed effects can be recovered from the estimated firm fixed effects and the individual means. The individual mean contains the average firm effect of the firms for which individual i worked, weighted by match length. The vector of these averages is equal to $T^{-1}M\hat{\Psi}^{ME}$, so that the OLS estimate of the individual fixed effects is:

$$\begin{aligned}
\bar{y}_i &= \bar{X}_i\hat{\beta}^{ME} + \hat{\theta}^{ME} + T^{-1}M\hat{\Psi}^{ME} \Leftrightarrow \\
\hat{\theta}^{ME} &= \bar{y}_i - \bar{X}_i\hat{\beta}^{ME} - T^{-1}M\hat{\Psi}^{ME} \Leftrightarrow \\
\hat{\theta}^{ME} &= \bar{\tilde{y}}_i - T^{-1}M\hat{\Psi}^{ME}
\end{aligned} \tag{6}$$

Finally, the match effects can be computed from the match means:

$$\begin{aligned}
\bar{y}_s &= \bar{X}_s\hat{\beta}^{ME} + \hat{\theta}_i^{ME} + \hat{\Psi}_j^{ME} + \hat{\lambda}_s^{ME} \Leftrightarrow \\
\hat{\lambda}_s^{ME} &= \bar{y}_s - \bar{X}_s\hat{\beta}^{ME} - \hat{\theta}_i^{ME} - \hat{\Psi}_j^{ME} \Leftrightarrow \\
\hat{\lambda}_s^{ME} &= \bar{\tilde{y}}_s - \hat{\theta}_i^{ME} - \hat{\Psi}_j^{ME}
\end{aligned} \tag{7}$$

Where \bar{y}_s and $\bar{\tilde{y}}$ are $S \times 1$ vectors containing the match means of y and $y - X\hat{\beta}^{ME}$ and \bar{X}_s is the $S \times K$ matrix containing the match means of X . Both calculations are computationally trivial.

The same algorithm can be applied to the TWFE model with one modification: In the TWFE model, the OLS predictions are different from the match means even if there really are no match effects,⁵ which makes the first step more complicated: One needs to run a partial regression by regressing y and each column of X on the individual and firm fixed effects and use the residuals to obtain $\hat{\beta}^{TW}$, the coefficient on X in the TWFE model. Implementation is simplified by the fact that these regressions have the same covariates (the two sets of fixed effects only) as the regression solved by equation 4. Consequently, one can use the same simplification as above by solving equation 5 and 6 repeatedly for each column of X in place of y . This modified first step is a simple way to implement the transformation proposed by Wansbeek and Kapteyn (1989) and Davis (2002). Rather than inverting the whole $(N + J) \times (N + J)$ matrix, it only requires computation of the inverse of $(F - K'T^{-1}K)$. This matrix is later used to compute the firm fixed effects, so it only needs to be computed once and can be reused in the auxiliary regressions for y and all elements of X as well as to obtain the estimates of the firm fixed effects from \tilde{y} after $\hat{\beta}^{TW}$ has been calculated. This advantage does not apply when using the CGA to solve equation 5, which has to be repeated for each covariate. In most cases, however, solving equation 5 with the CGA is a matter of seconds.

After this step, estimation proceeds exactly as in the match effects model: One obtains estimates of $\hat{\beta}^{TW}$ from a partial regression, subtracts the fitted values from y to obtain \tilde{y} and obtain the firm fixed effects either using the CGA or the inverse of $(F - K'T^{-1}K)$ from the partial regressions to solve equation 5. The individual effects are then given by equation 6.

Estimating variance matrices

⁵ To see this, note that the projection matrix of the match effect model computes the match means, which is not the case for the projection matrix of the TWFE model, $[D_\theta \ D_\psi][[D_\theta \ D_\psi]'[D_\theta \ D_\psi]]^{-1}[D_\theta \ D_\psi]'$.

The common strategy to simplify computation by iteratively solving the normal equations complicates inference, because it avoids inverting the cross-product of the data matrix, which is a key component of asymptotic formulas for the variance matrix.⁶ Using the bootstrap amplifies the computational problem and is additionally complicated by the necessity to maintain a connected sample (see Woodcock 2015 Appendix B for subsampling connected groups), especially when errors are correlated within firm or individual (Cameron, Gelbach and Miller 2011). The algorithm here yields the block of the inverse matrix corresponding to $\hat{\beta}$, so the asymptotic variance matrix is simple to compute regardless of whether ε is correlated across observations or not. SEs are frequently calculated under the assumption of uncorrelated errors, even though it is frequently violated in panel data and can seriously distort rejection rates (Kezdi 2004). In most applications, errors are likely correlated within both units for which fixed effects are included. Two-way clustered standard errors of the slopes are simple to obtain based on the covariance matrix from the partial regression to obtain $\hat{\beta}$ above. See Appendix B for an example.

Thereby, the algorithm facilitates specification tests that appear of key relevance in common applications, because misspecification in these models is just as problematic as in standard regressions. The TWFE is a constrained match effects model that restricts all match effects to equal zero, which causes bias if the restrictions do not hold (Woodcock 2008). The restrictions can be examined by testing whether the slope coefficients change (which requires estimating their variance matrix) or by comparing the restricted TWFE to the unrestricted match effects model (which requires estimating both models). A likely reason why such tests are rarely conducted is that it is often impossible or prohibitively costly to compute these tests. The method here estimates the variance matrix of the slopes and simplifies estimating the unrestricted model, thereby making both tests feasible.

Key Advantages and Comparison to Other Methods

⁶ See Cattaneo, Jansson and Newey (2018a) for a discussion of whether standard asymptotics apply. However, the alternative asymptotic formulas also require the (upper block) of $([X D_\theta D_\psi]'[X D_\theta D_\psi])^{-1}$ (see e.g. Cattaneo, Jansson and Newey, 2016a,b), so the problems discussed below apply in the same way.

In summary, the algorithm I propose facilitates estimating the fixed effects and specification tests. In addition, it reduces the computational burden by requiring less memory and computation time. The memory and time requirements and how they compare to other algorithms are case specific and depend heavily on the implementation.⁷ In theory, the factor dominating computational complexity is the size of the problem to be solved. For iterative solutions, both memory required and computation time grow (at least) by the square of the size of the problem. For direct solutions via Matrix inversion, computation time is of cubic order. Thus, the key simplification of the algorithm I propose stems from reducing the problem to size J . To illustrate the computational advantages of the approach above for a specific case (despite the caveats on comparability), Table 1 reports the computation times of different algorithms as the number of individuals and firms increase. The simulation setup is provided in Appendix C. For comparability, I focus on Stata implementations of the within transformation, the CGA and alternating projections.⁸

Table 1: Computation Time (in Minutes) of Different Stata Algorithms Using Simulated Data

		(1)	(2)	(3)	(4)	(5)	(6)
<i>Individuals</i>	<i>Firms</i>	<i>Match Effects</i>	<i>TWFE</i>	<i>felsdvreg</i>	<i>a2reg</i>	<i>reg2hdfe</i>	<i>reghdfe</i>
100,000	1000	0.2	0.2	0.6	0.5	29.1	0.6
1,000,000	10,000	2	2	40	5	302	7
10,000,000	100,000	19	26	-	51	-	73
<i>Program calculates...</i>							
Fixed Effects		Yes	Yes	Yes	No*	No*	No*
SEs of slopes		Yes	Yes	Yes	No	Yes	Yes
SEs of Fixed Effects		No	No	Firms only	No	No	No

Note: Column 1 reports computation time from estimating the match effects model using the algorithm described here, the remaining columns estimate the TWFE. Column 2 uses the method in this paper. Column 3 uses the within transformation described in Cornelissen (2008), column 4 uses the CGA (Abowd, Creedy and Kramarz, 2002) implemented by Ouazad (2007). Columns 5 and 6 use the alternating projections methods by Guimarães and Portugal (2010) and Correia (2014). All times are the average of 5 iterations in minutes. See appendix C for details on the simulation setup. There are 5 covariates and 10 observations per individual. Individuals move to a randomly chosen firm either once (with probability 0.16) or twice (with probability 0.04). All simulations were run single-threaded on a computer with 74 GB of main memory. * indicates that the program optionally calculates the statistic of interest, but it was not done in the simulation and may slightly increase computation time.

The within transformation proposed by Andrews, Schank and Upward (2006) implemented by Cornelissen (2008) solves a problem of size $K + J$. I additionally exploit the simple structure of the matrix and solve

⁷ Programs (twfe) for Matlab and Stata are available from my website and through SSC.

⁸ Differences in the underlying software, memory requirements and the practicability make comparisons to the Matlab, R and Fortran program even less likely to be meaningful.

the smaller problem by the CGA rather than directly. Column 3 of Table 1 shows that these additions speeds up computation dramatically, particularly as the size of the problem grows. The within transformation algorithm does not work for large problems, because the within transformation is more memory intensive than iterative solutions. The CGA approach by Abowd, Creedy and Kramarz (2002) implemented in Ouazad (2007) solves a problem of size $K + N + J$. Even in the most unfavorable case, my approach cuts the size of the problem in half. The results in column 4 show that this cuts computation time by a factor of two to three. The CGA scales well, so both approaches can solve large problems. Columns 5 and 6 report times from two implementations of the alternating projections approach by Guimarães and Portugal (2010) and Correia (2014). Arcidiacono et al. (2012) and Gaure (2013) also propose solutions based on alternating projections, all of which iteratively solve a problem of size $K + N + J$. The implementation by Guimarães and Portugal (2010) becomes very slow as the size of the problem grows. The program by Correia (2014) scales better in terms of computation time. An advantage of these solutions is their ease of implementation and that they are easier to extend to more than two sets of fixed effects and non-linear models.

4. Application to Wage Determinants and Matching in Germany

I use the algorithm introduced above to examine wage determinants using linked employer-employee data from Germany (LIAB mover model 9308). The data cover 1993 to 2008 and are designed to estimate models with individual and firm fixed effects. They only include firms that employed at least one worker who also worked for another firm in the panel, i.e. only firms for which the firm effect is identified. They contain all workers who moved between firms in the panel. My final sample contains 9,891,519 observations from 3,068,373 individuals working at 24,323 firms.⁹ Appendix D provides further

⁹ I follow the common practice to restrict the sample to the largest connected group, which contains 99% of all observations. I exclude individuals from the former GDR (East Germany) who entered the labor market before 1990, because their records may not be complete. All firms in the sample are connected to at least one other firm, but some firms form small isolated groups that are not connected to the largest group I use. In addition, some links may be broken by sample restrictions.

information on the data, potential issues such as topcoding as well as the sample I use. Appendix E reports summary statistics. I focus on key results that illustrate advantages of the algorithm I propose, mainly documenting that the inclusion of match effects affects estimates of key parameters and underlining the value of estimating and analyzing the fixed effects. Full results of all models are in appendix F.

I first use the algorithm described above to estimate a rich wage regression. I regress the log of person i 's daily wage at firm j in time period t (w_{ijt}) on characteristics of person i (x_{ijt}^I), and firm j (x_{ijt}^F) as well as year, firm, individual and match fixed effects:

$$\log(w_{ijt}) = \theta_i + \Psi_j + \lambda_s + \phi_t + x_{ijt}^F \beta^F + x_{ijt}^I \beta^I + \varepsilon_{ijt} \quad (8)$$

To analyze the effect of omitting the match effects, I also estimate the TWFE model, which omits λ_s from equation 8. I allow for two-way clustering, i.e. errors can be arbitrarily correlated within firms and individuals. I then use estimated fixed effects in the following regressions:

$$\begin{aligned} \hat{\Psi}_j &= z_j^F \delta + \eta_j^F \\ \hat{\theta}_i &= z_i^I \gamma + \eta_i^I \end{aligned} \quad (9)$$

where z_i^I are time invariant characteristics of firms and individuals. Such models can be used to examine which time-invariant characteristics of firms (z_j^F) and individuals (z_i^I) predict that a firm pays or an individual receives high wages.

Table 2 reports selected coefficients from the wage regressions defined by equation 8 and 9. The first question of interest is whether match effects matter and if so, whether specification tests detect them. In line with Card, Heining and Kline (2013), the variance of the match effects is small relative to the variance of the individual and firm fixed effects. Yet, the difference in the explained sum of squares in Table A5 shows that match effects explain more than 25 percent of the unexplained wage dispersion in Germany. In addition, including match effects appears to be important for the estimates of the slope coefficients. A Hausman-type test for equality of the slope coefficients between the two models clearly rejects the TWFE

(p-value of .00).¹⁰ Conducting this test requires the variance matrices of the slope coefficients, thereby emphasizing an advantage of the algorithm proposed here that automatically estimates this variance matrix.

Table 2: Determinants of log of Daily Wages, Selected Coefficients

	Match Effects	TWFE
Panel A: Fixed-effects Wage Regression		
Part time job	-0.3395*** (0.0046)	-0.4446*** (0.0059)
Number of years in current establishment	0.0099*** (0.0017)	0.0066*** (0.0002)
Experience in years	0.1579*** (0.0026)	0.1171*** (0.0014)
Panel B: Regressions of Individual FE on Time Invariant Individual Characteristics		
Female	-0.0559*** (0.0032)	-0.1035*** (0.0029)
<i>School education and vocational training</i>		
Secondary/intermediate school w/o completed vocational training	-0.2325*** (0.0042)	-0.2605*** (0.0038)
Upper secondary school w/o completed vocational training	-0.3391*** (0.0152)	-0.3914*** (0.0136)
Upper secondary school with completed vocational training	0.2338*** (0.007)	0.1849*** (0.0065)
Completion of a university of applied sciences	0.3861*** (0.007)	0.3147*** (0.0064)
College / university degree	0.5251*** (0.006)	0.4205*** (0.0053)
Missing	-0.0248*** (0.0061)	-0.1286*** (0.0056)
Notes: Number of observations: 9,792,405 (Panel A) and 3,062,118 (Panel B). Panel A is based on estimating equation 9, Panel B is based on estimating equation 10. Only selected coefficients are reported, see the appendix tables for full results and information on omitted categories. Table A5 in Appendix F reports all coefficients of the regression in Panel A. Table A6 contains coefficients from regressing firm fixed effects on firm characteristics (not reported in this table). Table A7 reports all coefficients from the regressions of individual fixed effects on time invariant characteristics in panel B. Standard errors are clustered at the firm and individual level. *: Significant at 5%; **: Significant at 1%; ***: Significant at 0.1%.		

An important question is whether the omission of match effects leads to wrong conclusions. Comparing the results from the TWFE and the match effects model in Table 2 reveals substantive differences: The returns to experience and tenure are flatter in the TWFE.¹¹ Part of the large difference between part- and full-time jobs in the TWFE is also due to the omission of match effects. The consequences of omitting the

¹⁰ This test can only detect match effects that are correlated with the covariates. Alternatively, one can conduct an F-test using the residual sum of squares, but this test is only valid under the assumption of uncorrelated errors.

¹¹ The regression includes higher order terms, which are not reported in Table 2, but do not affect this conclusion.

match effects follow from standard omitted variable bias formulas, which simplify because the normalizations of the match effects make some terms mechanically equal zero. As Woodcock (2008, p.779 especially equation 16) points out, the effect of omitting the match effects on the estimated slope coefficients arises from the conditional covariance between X and D_λ , given D_θ and D_ψ . That is, the bias arises from match dummies being more predictive of the included covariates than firm and individual dummies only. For example, part-time jobs tend to be worse matches than full-time jobs, even when only comparing jobs within individual and firm (i.e. conditional on D_θ and D_ψ). Thus, failing to control for match quality overstates the wage difference between part- and full-time jobs.

The lower part of Table 2 shows that omitting match effects not only affects the estimated slopes, but also the fixed effects and hence parameters calculated from them. Most importantly, the estimated gender wage gap in the TWFE is almost twice as large as the 5% gap in the match effects model. The education gradient is steeper in the match effects model. Individual and match fixed effects are orthogonal by construction, so the bias arises only because omitted variable bias spreads from time-varying characteristics (and possibly bias in the firm fixed effects).¹² For example, part time jobs are more likely to be poor matches. Hence, excluding match effects biases the estimated effect of part time jobs downward. According to the standard omitted variable bias mechanics, part of this bias spreads to all other covariates that are correlated with the part time status. Females are more likely to work in part time jobs, i.e. the covariance between the two variables is positive. Both variables are negatively correlated with wages. Therefore, one would expect this channel to lead to negative bias in the gender wage gap. In

¹² To see this, consider the second line of equation 16 in Woodcock (2008): Without covariates in equation 10, the bias in the gender wage gap would be given by the gender difference in the average bias in $\hat{\theta}$, i.e. how the average of $(D'_\theta M_{[X D_\psi]} D_\theta)^{-1} D'_\theta M_{[X D_\psi]} D_\lambda \lambda$ differs by gender, where $M_{[X D_\psi]}$ is the matrix that projects out X and D_ψ . Match and individual dummies are orthogonal ($D'_\theta D_\lambda = 0$), but not conditional on X and firm dummies, i.e. $D'_\theta M_{[X D_\psi]} D_\lambda$ can be non-zero. This term weights the match effects by a term that depends on gender differences in the covariates. These weights may differ between jobs due to time-varying characteristics, which explains why the bias can be non-zero even though match effects sum to zero for each person and hence for both males and females. Covariates in equation 10 makes this problem more complex.

line with this logic, the overall bias in the gender wage gap is negative. Even though it is an indirect effect, the bias is substantial, which emphasizes that even if one is only interested in the fixed effects, one should not be casual about the choice of regressors and the fixed effect structure.

In addition to the slope coefficients, estimates of the fixed effects are of interest for a wide range of topics including wage differentials (e.g. Woodcock 2015), wage dynamics (e.g. Abowd, Kramarz and Roux 2006), inequality (e.g. Card, Heining and Kline 2013), teacher value added (e.g. Rockoff 2004) and teacher mobility (Jackson 2013). In order to illustrate that the fixed effects produced by the method introduced in this paper can be used to shed light on the importance of various mechanisms, I analyze how pre-match characteristics affect the permanent wage components of the subsequent job, i.e. the firm and match fixed effect. Estimating the fixed effects and relating them to pre-match characteristics, such as whether the match was preceded by an unemployment spell, allows me to assess whether these effects are partly due to the conditions when matches are formed. Permanent wage effects of the conditions under which matches are formed have been documented for, among others, employment after mass layoffs (e.g. Jacobson, LaLonde and Sullivan 1993), economic conditions during job search (e.g. Oreopolous, von Wachter and Heisz 2012, Schmieder, von Wachter and Heining 2018) and age of the employee (e.g. Topel and Ward 1992). Most previous papers restricted attention to jobs after a particular event, such as mass layoffs, that affect only a small fraction of individuals in order to isolate the effect of a specific factor (e.g. involuntary unemployment in case of mass layoffs) on subsequent wages.

Being able to estimate both firm and match fixed effects for a large sample of movers, I look at these permanent wage effects from a different angle. Rather than restricting the sample to specific job transitions (e.g., transitions following a mass layoff), I use all individuals with two or more jobs and exploit the rich information in the IAB data to measure the circumstances before employment. To do so, I regress match effects and the corresponding firm effects on pre-match characteristics:

$$\begin{aligned}\hat{\lambda}_s &= z_s^M \pi_1 + \eta_s^M \\ \hat{\Psi}_s &= z_s^M \pi_2 + v_s^M\end{aligned}\tag{10}$$

where $\hat{\Psi}_s$ is the estimated firm fixed effect of the firm that match s corresponds to and z_s^M is a vector of variables measured at the time match s is formed.

The sum of the two effects is the expected difference between the wage of the job an individual accepted and the wage at a randomly assigned job by virtue of both match and firm effects being normalized to sum to zero. Thus, the results provide evidence on the circumstances that lead to matches with permanently high or low wages. The results are in table 3:¹³ The first column provides evidence on the conditions under which job transitions lead to good matches between worker and firm, i.e. wages that are unusually high for both the worker and the firm. The second column examines which characteristics lead to a worker being matched with a high-wage firm, i.e. a firm that pays high wages to all employees. Involuntary job transitions have long lasting negative effects, but job transitions are also an important source of wage growth (Topel and Ward 1992). The results in table 3 clarify that permanent wage changes at job transitions are primarily due to moving to higher paying firms: The small coefficients and the low R^2 in column 1 of Table 3 show that the relation between match quality and observable pre-match characteristics is weak compared to firm effects. Consequently, firm heterogeneity not only plays an important role in wage dispersion (Gruetter and Lalive 2009), rising inequality (Card, Heining and Kline 2013), the effects of unemployment insurance (Nekoei and Weber 2017, Schmieder, von Wachter and Heining 2018), but is also a key part of wage dynamics over the life-cycle. The coefficients on the dummies for the type of transition (job-to-job, non-employment-to-job, training-to-job and initial job)¹⁴ in table 3 suggest that this importance of firm heterogeneity not only applies to job-to-job transitions as

¹³ The regression only uses individuals for whom non-zero match effects are identified, i.e. movers. I do not consider job transitions within the same firm, since they are likely to be different in terms of job search and signaling.

¹⁴ Transition type is defined using labor market status 8 days before the current match to avoid misclassifying job-to-job transitions with a short break.

Haltiwanger, Hyatt and McEntarfer (2018) document. The omitted category is job-to-job transitions, so as most search models would predict, initial employment is a slightly worse match. All other wage effects of the type and duration of the previous labor market state operate through the firm fixed effect: initial employment and employment after an episode without employment is at lower paying firms than job-to-job transitions (by 8 and 6 percentage points). These results suggest that the “job ladder” also plays a key role for other transition types and that workers may travel it in both directions.

Table 3: Regression of Match and Firm FE on Pre-Match Characteristics, Selected Coefficients

	Match Effect	Firm Effect
<i>Employment status 8 days before current match</i>		
Previous spell was benefits or gap	-0.0012 (0.0039)	-0.057*** (0.0107)
Apprentice/trainee at other Firm	0.0074 (0.0155)	-0.0164 (0.0488)
No previous record	-0.04*** (0.0066)	-0.0796*** (0.0176)
Number of days in labor market status 8 days before current match (main effect)	-0.000001 (0.000001)	0.000009*** (0.000002)
...if previous spell was benefits or gap (interaction)	0.000001 (0.000003)	0.000003 (0.000008)
...if previous spell was training (interaction)	0.000016 (0.000019)	0.000009 (0.000058)
Number of years of benefit receipt up to beginning of current match	0.0027* (0.0012)	-0.0301*** (0.0032)
Part time job (at beginning of match)	-0.1187*** (0.0047)	-0.6257*** (0.0151)
R^2	0.0343	0.2189

Note: 665,080 observations. The omitted employment status before the current match is “employment at a different firm”. Only selected coefficients are reported, Table A8 in Appendix F reports all coefficients. *: Significant at 5%; **: Significant at 1%; ***: Significant at 0.1%.

4. Conclusion

Models with two or more large sets of fixed effects are heavily used in a wide range of fields such as labor economics, education, health and migration. Their advantages are widely recognized, but their use is limited by computational difficulties. I propose a simple method to compute the OLS estimates of the TWFE and the match effects model in large data sets that substantially reduces computational complexity. It not only offers advantages in terms of speed and computational resources needed, but also simplifies estimation of the fixed effects and the variance matrix of the slopes that can allow for multi-way clustering. Using the algorithm to analyze wage determinants in Germany underlines key advantages: Important

parameter estimates, such as the gender wage gap, change substantially when including match effects. The algorithm I propose facilitates detecting and formally testing for such changes. The application underlines the usefulness of computing the fixed effects, because analyzing them can provide insights on economic questions regardless of their status as structural parameters. In particular, firm fixed effects play a key role in determining the gains and losses from job transitions, emphasizing the importance of firms in wage dynamics.

References

- Abowd, J. M., F. Kramarz, and D. N. Margolis, 1999, High wage workers and high wage firms. *Econometrica* 67 (2), 251-334.
- Abowd, J. M., R. H. Creecy, and F. Kramarz, 2002. Computing person and firm effects using linked longitudinal employer-employee data. Census Bureau Technical Paper TP-2002-06.
- Abowd, J. M., F. Kramarz, and S. Roux, 2006, Wages, Mobility and Firm Performance: Advantages and Insights from Using Matched Worker-Firm Data. *The Economic Journal*, 116(512), F245–F285.
- Abowd, J.M., F. Kramarz and S.D. Woodcock, 2008, Econometric Analyses of Linked Employer-Employee Data, in: L. Mátyás, P. Sevestre (Eds.), *The Econometrics of Panel Data*, Chapter 22. Springer, Heidelberg, pp. 727-759.
- Andrews, M., T. Schank and R. Upward, 2006. Practical fixed-effects estimation methods for the three-way error-components model. *Stata Journal*, 6(4), 461-481.
- Arcidiacono, P., G. Foster, N. Goodpaster, and J. Kinsler, 2012. Estimating Spillovers using Panel Data, with an Application to the Classroom. *Quantitative Economics*. 3(3), 421–470.
- Bennett, D., C. L. Hung, and T. L. Lauderdale, 2015. Health care competition and antibiotic use in Taiwan. *The Journal of Industrial Economics*, 63(2), 371-393.
- Cameron, C. A., J. B. Gelbach and D.L. Miller, 2011, Robust Inference With Multi-Way Clustering. *Journal of Business and Economics Statistics* 29 (2), 238-249.
- Card, D., J. Heining, and P. Kline, 2013. Workplace Heterogeneity and the Rise of West German Wage Inequality. *The Quarterly Journal of Economics*, 128(3), 967-1015.
- Cattaneo, M. D., M. Jansson, and W. K. Newey, 2018a. Inference in Linear Regression Models with Many Covariates and Heteroskedasticity. *Journal of the American Statistical Association*, 113(523), 1350-1361.
- Cattaneo, M. D., M. Jansson, and W. K. Newey, 2018b. Alternative Asymptotics and the Partially Linear Model with Many Regressors. *Econometric Theory*, 34(2), 277-301.
- Cornelissen, T., 2008. Using Stata for a memory-saving fixed-effects estimation of the three-way error-components model. *Stata Users Group German Stata Users' Group Meetings* 2008 07.
- Correia, S., 2014. "REGHDFE: Stata module to perform linear or instrumental-variable regression absorbing any number of high-dimensional fixed effects," *Statistical Software Components S457874*, Boston College Department of Economics, revised 25 Jul 2015.
- Davis, P. 2002. Estimating multi-way error components models with unbalanced data structures. *Journal of Econometrics* 106, 67-95.
- Fischer, G., F. Janik, D. Müller and A. Schmucker 2008, The IAB establishment panel - from sample to survey to projection. *FDZ Methodenreport* 01/2008 (en).

- Garen, J.E. 1989. Job-Match Quality as an Error Component and the Wage-Tenure Profile: A Comparison and Test of Alternative Estimators. *Journal of Business & Economic Statistics*, 7(2), 245-252.
- Gaure, S., 2013. OLS with multiple high dimensional category variables. *Computational Statistics & Data Analysis*, 66, 8–18.
- Greene, W.H. 2008, *Econometric Analysis*, 6th Edition. Prentice Hall, New York.
- Grogger, J. and G.H. Hanson, 2011, Income maximization and the selection and sorting of international migrants. *Journal of Development Economics* 95, 42-57.
- Gruetter, M. and R. Lalive, 2009. The importance of firms in wage determination. *Labour Economics*, 16(2), 149–160.
- Guimarães, P. and P. Portugal, 2010. A Simple Feasible Alternative Procedure to Estimate Models with High-Dimensional Fixed Effects. *Stata Journal*, 10(4), 628-649.
- Haltiwanger, J., H. Hyatt and E. McEntarfer, 2018. Who Moves Up the Job Ladder? *Journal of Labor Economics* 36(S1), S301-S336.
- Jackson, C.K. 2013, Match Quality, Worker Productivity, and Worker Mobility: Direct Evidence from Teachers. *Review of Economics and Statistics*, 95(4), 1096–1116.
- Jacob, B.A., and L. Lefgren, 2008. Can Principals Identify Effective Teachers? Evidence on Subjective Performance Evaluation in Education. *Journal of Labor Economics*, 26(1), 101–136.
- Jacobebbinghaus, P. 2008, LIAB-Datenhandbuch, Version 3.0. FDZ Datenreport, 03/2008 (de).
- Jacobson, L. S., R. J. LaLonde, and D. G. Sullivan, 1993. Earnings Losses of Displaced Workers. *American Economic Review*, 83(4), 685–709.
- Jovanovic, B. 1979. Firm-specific Capital and Turnover. *Journal of Political Economy* 87(6), 1246-1260.
- Kezdi, G. 2004, Robust Standard Error Estimation in Fixed-Effect Panel Models. *Hungarian Statistical Review. Special English Volume* 9, 95-116.
- Koch, I. and H. Meinken 2004. The Employment Panel of the German Federal Employment Agency. *Journal of Applied Social Science Studies* 124(2), 315-325.
- Kramarz, F., S. Machin and A. Ouazad 2008. What Makes a Test Score? The Respective Contributions of Pupils, Schools, and Peers in Achievement in English Primary Education. IZA Discussion Paper 3866.
- Le Gall, F., 2014. Powers of Tensors and Fast Matrix Multiplication, 1–28. Retrieved from <http://arxiv.org/abs/1401.7714v1>
- Mortensen, D.T. 1978. Specific Capital and Turnover. *The Bell Journal of Economics* 9(2), 572-586.
- Nekoei, A. and A. Weber, 2017. Does Extending Unemployment Benefits Improve Job Quality? *American Economic Review* 107(2), 527–561.
- Ouazad, A., 2007. Program for the Estimation of Two-Way Fixed Effects, available at <http://personal.lse.ac.uk/ouazad/>.

- Oreopoulos, P., T. von Wachter and A. Heisz, 2012. The Short- and Long-Term Career Effects of Graduating in a Recession. *American Economic Journal: Applied Economics*, 4(1), 1–29.
- Rockoff, J. E. 2004. The Impact of Individual Teachers on Student Achievement: Evidence from Panel Data. *American Economic Review*, 94(2), 247–252.
- Schmieder, J.F., T. von Wachter and J. Heining, 2018. The Costs of Job Displacement over the Business Cycle and Its Sources: Evidence from Germany. Working paper.
- Theil, H. 1971, *Principles of Econometrics*. John Wiley and Sons, New York.
- Topel, R.H. and M.P. Ward 1992, Job Mobility and the Careers of Young Men. *The Quarterly Journal of Economics* 107(2), 439-479.
- Wansbeek, T. and A. Kapteyn 1989, Estimation of the Error-Components Model with Incomplete Panels. *Journal of Econometrics* 41, 341-361.
- Woodcock, S.D, 2008, Wage Differentials in the presence of unobserved worker, firm and match heterogeneity. *Labour Economics* 15, 772-794.
- Woodcock, S.D. 2015, Match Effects. *Research in Economics*, 69(1), 100-121
- Yule, G.U. 1907, On the Theory of Correlation for any Number of Variables, Treated by a New System of Notation. *Proceedings of the Royal Society of London Series A, Containing Papers of a Mathematical and Physical Character* 79 (529), 182-193.