COHERE - Centre of Health Economics Research, Department of Business and Economics
Discussion Papers, No. 2013:6
ISSN: 2246-3097

A simple but efficient approach to the analysis of multilevel data

By
Stefan Holst Milton Bache (*), Troels Kristensen (**)
*COHERE, Department of Business and Economics, University of Southern Denmark
**COHERE, Institute of Public Health, Research Unit of General Practice, and University of Southern Denmark

COHERE, Department of Business and Economics
Faculty of Business and Social Sciences
University of Southern Denmark
Campusvej 55
DK-5230 Odense M
Denmark
www.cohere.dk

A simple but efficient approach to the analysis of multilevel data

Stefan Holst Milton Bache *[1,2,5] and Troels Kristensen[1,3,4,5]

[1]Centre of Health Economics Research, COHERE
[2]Department of Business and Economics
[3]Institute of Public Health
[4]Research Unit of General Practice
[5]University of Southern Denmark

May 2013.
Words ≈ 4,200.

Abstract

Much research in health economics revolves around the analysis of hierarchically structured data. For instance, combining characteristics of patients with information pertaining to the general practice (GP) clinic providing treatment is called for in order to investigate important features of the underlying nested structure. In this paper we offer a new treatment of the two-level random-intercept model and state equivalence results for specific estimators, including popular two-step estimators. We show that a certain encompassing regression equation, based on a Mundlak-type specification, provides a surprisingly simple approach to efficient estimation and a straightforward way to assess the assumptions required. As an illustration, we combine unique information on the morbidity of Danish type 2 diabetes patients with information about GP clinics to investigate the association with fee-for-service healthcare expenditure. Our approach allows us to conclude that explanatory power is mainly provided by patient information and patient mix, whereas (possibly unobserved) clinic characteristics seem to play a minor role.

Keywords: Multilevel models, random intercepts, nested models, Mundlak device, correlated random effects, 2-step estimation, estimated dependent variables, fee-for-service expenditures, type 2 diabetes.
Running head: Simple and efficient analysis of multilevel data.

*Corresponding author. Campusvej 55, DK-5230 Odense M, Denmark. Email: stefan@sdu.dk. Phone: (+45) 6550 3887 / (+45) 3113 1369.

1 Introduction

The rapidly increasing richness of available data and the ability to link and merge these across various unit-specific levels have brought much recent attention to statistical multilevel models. These methods acknowledge and seek to utilize the nested data structure. Rice and Jones (1997) provide an introductory account and point to areas within health economics where these methods could prove beneficial. A few other field-related examples are Laudicella et al. (2010); Scribner et al. (2009); Fletcher (2010); Gurka et al. (2011); Carey (2000); and Blundell and Windmeijer (1997).

While researchers can choose among complex models for multilevel analysis, simpler models have the benefit of practicality, ease of interpretation, and fewer distributional assumptions. Combining deep hierarchies with random slopes and cross-level interactions can be challenging and may not be necessary to answer the question at hand. At the price of some composition detail, simpler regression designs offer robustness and practicality by weakening distributional requirements and allowing the use of a more conventional regression framework. It is important to strike a balance while acknowledging the possible importance of the underlying nested data structure for a particular problem.

In this paper we present a new treatment of the simple, yet very useful, two-level random intercepts model. We provide insights that relate various estimation strategies and their associated parameter estimators. From a practitioner's view, our results will clarify what is being estimated at the two levels and which assumptions are required to answer a particular question. We argue that a combined equation encompassing both levels is useful in several aspects. In particular, we show that an elegant result by Mundlak (1978), which bridges *fixed effects* and *random effects* estimation, carries over to the current multilevel setting.[1]

The remainder of the paper is structured as follows. In Section 2, we discuss the model setup, estimation strategies, and our main results. In Section 3 we illustrate the results by applying the model to unique data on the co-morbidity of type 2 diabetes patients combined with clinic information. Our aim is to investigate whether the available information is able to explain the variation in fee-for-service expenditure and how explanatory power and unexplained variation is distributed at the two levels. After giving concluding remarks in Section 4, we provide verification of the results in Appendix A and some detail on the practical implementation in Appendix B.

2 The model and estimator relationships

Suppose we are faced with a research question revolving around individuals each of whom belongs to one of several groups. Here we use the terms "individuals" and "groups", but this could be any relevant nesting of units. Examples are patients nested within GP clinics or hospital departments; doctors nested within hospitals; or certain operations performed in different operating theatres.

We consider an individual-level equation specified as

$$y_i = \boldsymbol{x}_i'\boldsymbol{\beta} + \gamma_{j(i)} + u_i, \tag{1}$$

[1] Unfortunately, our terminology can lead to confusion: the terms "fixed effects" and "random effects" here take their meaning from panel data econometrics and should not be confused with their use in multilevel models.

where y_i is the outcome variable to be explained; $\boldsymbol{x}_i$ is a vector of individual characteristics; $\gamma_{j(i)}$ is an (unobserved) component, or *effect*, specific to group j of which individual i is a member; and u_i is the unexplained noise term. We use boldface to denote vectors (which we take as columns), and the prime ($'$) denotes transposition.

At the second level of the hierarchy, we specify the group effects γ_j in terms of group-level characteristics $\boldsymbol{z}_j$; in particular we let

$$\gamma_j = \boldsymbol{z}_j'\boldsymbol{\alpha} + e_j, \tag{2}$$

where the vector $\boldsymbol{z}_j$ is assumed to contain an intercept term. The vectors $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are unknown quantities of interest and are to be estimated. In addition to the parameter vectors, some quantification of the explained and unexplained variation at each level is typically also of direct interest.

The combined model is referred to as a *random intercepts* model, where the second-level equation specifies the intercepts in terms of the group-level data and a remainder. Surely, some applications may opt for models that also address the potential need for specifying random slopes and/or nesting structures deeper than two levels. We limit our discussion to the model specified by equations (1) and (2): it is common due to its simple interpretation, and it is practically straightforward to implement within a standard (least-squares) regression framework. Furthermore, under certain assumptions the slope parameters may be interpreted as averages of potentially random slopes which may suffice in the particular application.[2]

One approach that has been used for estimation of the model is to (separately)

[2] Typically, one needs some variation of the assumption that deviations from slope means should be uncorrelated with the included regressors.

estimate equation (1) using the *within* estimator (often also referred to as the *fixed effects* estimator). Estimation is based on the equation

$$y_i - \bar{y}_j = (\boldsymbol{x}_i - \bar{\boldsymbol{x}}_j)'\boldsymbol{\beta} + u_i - \bar{u}_j, \tag{3}$$

where group averages are subtracted from individual observations. Then, the unobserved effects γ_j are replaced by their estimates $\hat{\gamma}_j$ in equation (2) to enable estimation of the second-level equation; see e.g. Laudicella et al. (2010). The rationale behind this strategy is that elements with no variation within each group are eliminated by the *within transformation*, and one can thus disregard any potential dependence between $\boldsymbol{x}_i$ and $\gamma_{j(i)}$. However, the details of the second-stage estimation are less clear, and one may ask whether the problem solved in the first stage remains in the second. Furthermore, even though an estimated dependent variable need not invalidate a regression, there are varying numbers of observations contributing to each estimated group effect. One may therefore consider possible efficiency improvements associated with various weighted estimation procedures, see e.g. Lewis and Linzer (2005). It is also not clear at the outset how to decompose goodness-of-fit measures to asses the explanatory power of the information included, and to which degree estimation error from the first stage affects measures of variation.

An alternative to the two-stage strategy is to combine equations (1) and (2),

$$y_i = \boldsymbol{x}_i'\boldsymbol{\beta} + \boldsymbol{z}_{j(i)}'\boldsymbol{\alpha} + e_{j(i)} + u_i, \tag{4}$$

and employ (feasible) generalized least squares (GLS) for estimation. This is commonly referred to as (pure) *random effects* estimation. In this approach the variance structure of the combined error term is modeled explicitly. For consistent estimation of the model parameters, the strategy requires the unobserved effects (now $e_{j(i)}$) to be uncorrelated with the individual characteristics, $\boldsymbol{x}_i$, in addition to the group characteristics $\boldsymbol{z}_{j(i)}$; an assumption which may often be violated. On the other hand, when the assumption is satisfied the estimator can be more efficient than the two-step approach. The choice between strategies is at the core of the well-known "fixed effects versus random effects" discussion in panel data analysis.

We will argue that the so-called *Mundlak device* allows for a one-stage single-equation estimation where the choice between the two approaches is both arbitrary and unnecessary, to quote Mundlak (1978). Furthermore, our treatment makes the relation between the various estimators and weighting schemes more apparent. With a single equation, standard regression results make it clear which assumptions are needed. As an additional feature, a test for the appropriateness of the random effects estimation becomes directly available. The unification of approaches, in terms of the equivalence of estimators, is useful since the Mundlak-type equation is well-suited for estimation and offers some additional features, while the two-step approach provides intuition and ease of interpretation. Each approach could possibly offer extensions which are not straightforward for the other.

As argued above, a natural concern with estimation of (1) or the combined model in (4), treating γ_j (or e_j) as random, is whether exogeneity of $\boldsymbol{x}_i$ is reasonable. In this paper we focus on endogeneity due to the presence of $\gamma_{j(i)}$ and maintain the assumption that u_i is not a source of concern.[3] The *within* transformation in (3)

[3] If one suspects endogeneity due to elements in u_i, the current setup could be combined with methods that take this into account.

eliminates $\gamma_{j(i)}$ and thereby any endogeneity problems associated with this term. Using the resulting estimator, which we denote $\hat{\boldsymbol{\beta}}_w$, one can obtain estimates $\hat{\gamma}_{j,w}$ as the group means of the residuals $\hat{a}_i = y_i - \boldsymbol{x}_i'\hat{\boldsymbol{\beta}}_w$. On the other hand, correlation between $\gamma_{j(i)}$ and elements in $\boldsymbol{x}_i$ could be due to observable group characteristics, and using the specification in (2) could be sufficient in alleviating the endogeneity if $e_{j(i)}$ is uncorrelated with $\boldsymbol{x}_i$. In this case, estimation can be based on (4), possibly with specific assumptions on the covariance structure of the composite error term.

Now, if $e_{j(i)}$ is correlated with $\boldsymbol{x}_i$, one may attempt to capture the pertinent correlation with a linear projection

$$e_j = \bar{\boldsymbol{x}}_j'\boldsymbol{\pi} + v_j, \tag{5}$$

where the bar is used to denote group averages. We now plug this into the combined equation, which now reads

$$y_i = \boldsymbol{x}_i'\boldsymbol{\beta} + \boldsymbol{z}_{j(i)}'\boldsymbol{\alpha} + \bar{\boldsymbol{x}}_{j(i)}'\boldsymbol{\pi} + v_{j(i)} + u_i. \tag{6}$$

The averages often have a suitable interpretation as group-level aggregation of individual-level data, but in general this should be thought of as a technical device, often referred to as the *Mundlak device*. For reference, we explicitly write out the second-level equation for the groups as

$$\gamma_j = \boldsymbol{z}_j'\boldsymbol{\alpha} + \bar{\boldsymbol{x}}_j'\boldsymbol{\pi} + v_j. \tag{7}$$

It is natural to question the extent to which the projection can solve the endogeneity problem. It turns out that OLS estimation of (6) yields an estimator of $\boldsymbol{\beta}$ *identical* to $\hat{\beta}_w$, which we know to be robust to any kind of dependence between $e_{j(i)}$ and elements of $\boldsymbol{x}_i$.

We are typically also interested in the variance components of the composite error term, σ_u^2 and σ_v^2, say, which coupled with generalized least-squares (GLS) could also improve estimation efficiency. Suppose $\text{cor}(u_i, v_{j(i)}) = 0$, which leads to the classical random-effects covariance structure. Curiously, the GLS estimator of β based on (6) is also identical to $\hat{\beta}_w$. The "augmented" regression equation is therefore equivalent to the first stage in terms of estimating β regardless of whether OLS or GLS is used for estimation.[4]

We now discuss how the estimators of α are related. First, note that the uncertainty associated with the average member of a group is proportional to $1/n_j$, i.e. $\text{var}(\bar{u}_j) = \sigma_u^2/n_j$, if the u_i terms are uncorrelated. In the absence of group-specific noise, this would hint at the use of $\omega_j = n_j$ as regression weights when analyzing data aggregated at the group level. Since the precision with which each γ_j is estimated in the two-stage procedure is different for each j, it seems natural to assign higher weights to more informative groups in the second regression. If there is group-level noise, v_j, then aggregated uncertainty is $\text{var}(v_j + \bar{u}_j) = \sigma_v^2 + \sigma_u^2/n_j = \sigma_v^2(1 + (\sigma_u^2/\sigma_v^2)/n_j)$, which instead would suggest using second-stage weights $\widetilde{\omega}_j = 1/[1 + (\sigma_u^2/\sigma_v^2)/n_j]$. In fact, $\hat{\alpha}_{\text{OLS}}$, the estimator based on applying OLS to (6), is identical to the second-stage estimator based on (7) with ω_j as regression weights (and $\hat{\gamma}_{j,w}$ as dependent variable). Furthermore, $\hat{\alpha}_{\text{GLS}}$, the GLS estimator based on (6), is identical to second-stage regression with $\widetilde{\omega}_j$ as weights. While one may also consider an unweighted second-stage regression

[4] However, standard inference is not equivalent. OLS estimation, for example, should here be coupled with inference measures that are robust to clustering.

(which in terms of estimating $\boldsymbol{\alpha}$ is equivalent to a regression of (6) with weights $1/n_j$), it is advisable to gain efficiency by using appropriate weights, in particular if g is small.

An additional feature of the encompassing Mundlak-type equation relates to the parameter $\boldsymbol{\pi}$, which is zero if and only if $e_{j(i)}$ and $\boldsymbol{x}_i$ are uncorrelated. Testing the hypothesis $\boldsymbol{\pi} = \mathbf{0}$ therefore leads to a Hausman-type test for consistency of pure *random-effects* estimation based on (4). While the original form of such a test cannot be made robust to violations of the random-effects model assumptions, this version of the test can be made fully robust. For more details about this version of the test, see Baltagi (2006) and for its original form, see Hausman (1978). If $\boldsymbol{\pi} = \mathbf{0}$, equation (4) and a feasible GLS estimator should be used for estimation. It is also important to note from (7) that when $\boldsymbol{\pi} \neq \mathbf{0}$, a second-stage regression will be biased and inconsistent for $\boldsymbol{\alpha}$ if $\bar{\boldsymbol{x}}_j$ is excluded but correlated with $\boldsymbol{z}_j$.

We now summarize the main results of this paper:

(i) The estimators $\hat{\boldsymbol{\beta}}_w, \hat{\boldsymbol{\beta}}_{\text{GLS}}$, and $\hat{\boldsymbol{\beta}}_{\text{OLS}}$ are equivalent.

(ii) The estimators $\hat{\boldsymbol{\alpha}}_\omega$ and $\hat{\boldsymbol{\pi}}_\omega$ are equivalent to $\hat{\boldsymbol{\alpha}}_{\text{OLS}}$ and $\hat{\boldsymbol{\pi}}_{\text{OLS}}$, respectively.

(iii) The estimators $\hat{\boldsymbol{\alpha}}_{\widetilde{\omega}}$ and $\hat{\boldsymbol{\pi}}_{\widetilde{\omega}}$ are equivalent to $\hat{\boldsymbol{\alpha}}_{\text{GLS}}$ and $\hat{\boldsymbol{\pi}}_{\text{GLS}}$, respectively.

(iv) If $\boldsymbol{\pi} \neq \mathbf{0}$ and correlations exist among variables in $\boldsymbol{x}_i$ and $\boldsymbol{z}_j$, then the estimators of $\boldsymbol{\alpha}$ from (2) will be biased and inconsistent.

Here, $\hat{\boldsymbol{\alpha}}_\omega$, $\hat{\boldsymbol{\pi}}_\omega$, $\hat{\boldsymbol{\alpha}}_{\widetilde{\omega}}$, and $\hat{\boldsymbol{\pi}}_{\widetilde{\omega}}$ are the second-stage weighted least squares estimators based on (7) with ω, respectively $\widetilde{\omega}$, as regression weights and $\hat{\gamma}_{j,w}$ as dependent variable. For verification of the results, see Appendix A.

Part (i) is basically Mundlak's celebrated result, but here it is stated in an unbalanced setting (i.e. we have unequal group sizes), and we include group-specific explanatory variables.[5] A generalization of Mundlak's result with time-invariant variables is shown by Krishnakumar (2006) but also in a balanced panel data setting. Part (i) shows that the Mundlak device is sufficient for dealing with any kind of dependence between e_j and $\boldsymbol{x}_i$, since it yields an estimator of $\boldsymbol{\beta}$ that is identical to the one obtained by applying a within transformation (and because this is robust to any dependence structure). Therefore, since unbiased and consistent estimation of $\boldsymbol{\beta}$ using the within estimator requires $\mathrm{E}[u_i|\boldsymbol{x}_i, \bar{\boldsymbol{x}}_{j(i)}] = 0$, this is also the appropriate condition for the OLS/GLS estimator of (6). A relaxation of this assumption to zero correlation retains the consistency of the estimator(s).

Parts (ii) and (iii) relate one- and two-step estimators of $\boldsymbol{\alpha}$ and $\boldsymbol{\pi}$. Since we do not eliminate group-specific variables in estimation based on (6), we can directly obtain the estimates with no need for a second regression. On the other hand, the results justify a (weighted) second-stage regression where $\bar{\boldsymbol{x}}_j$ is included. To estimate $\boldsymbol{\alpha}$ and $\boldsymbol{\pi}$ based on (6) without bias, a sufficient assumption is $\mathrm{E}[u_i + v_{j(i)}|\boldsymbol{x}_i, \boldsymbol{z}_{j(i)}, \bar{\boldsymbol{x}}_{j(i)}] = 0$. Again, consistency only requires zero correlation. The equivalence results imply that the condition can be used for the second-stage estimators as well.

Part (iv) reveals that if there is correlation between $e_{j(i)}$ and $\boldsymbol{x}_i$ (the reason for employing the within estimator in the first place), then a second-stage regression based only on (2) is ill-advised if elements of $\boldsymbol{z}_{j(i)}$ are correlated with elements in $\boldsymbol{x}_i$. When separating the combined model into two regressions, and estimating

[5] Mundlak used a balanced panel data setting, where individuals are observed over time, and did not include time-invariant explanatory variables.

the first stage by eliminating the group-specific elements, one may unconsciously overlook the possible dependence issues in the second regression (which can be solved by the Mundlak correction). Even though one has controlled for correlation issues when estimating $\boldsymbol{\beta}$, any such problem often remains in the second-stage equation. Equations (6) and (7), where the Mundlak device is included, make it very clear when omission of $\bar{\boldsymbol{x}}_j$ leads to inconsistent estimation. It is also important to realize that while $e_{j(i)}$ is allowed to be correlated with $\boldsymbol{x}_i$, it is not allowed for v_j to be correlated with $\boldsymbol{z}_j$ (yet we could still consistently estimate $\boldsymbol{\beta}$). Finally, it is obvious that the within transformation—however powerful—does not eliminate problems that might arise from correlations between u_i and $\boldsymbol{x}_i$, and then nor does estimation based on (6).

To summarize, one can obtain the various two-step estimators directly from estimation based on (6), a Hausman-type test becomes directly available, and it is more apparent when estimation will be consistent and unbiased by applying the usual least squares regression reasoning. In practice, we do not know the variances σ_u^2 and σ_v^2, but it is advisable to employ a feasible (GLS) procedure to estimate (6) since using the weights ω_j may be inappropriate as seen from the expression of the variance term of the combined error. In Appendix B we review an approach to feasible GLS estimation. It should also be mentioned that using OLS on the second stage using (7) without weights is possible and provides consistent estimates under the usual assumptions, but since there are typically not many observations at this level, one would often try to get an efficiency gain by using proper weights. Finally, if one is worried about the assumed variance structure, it is possible to use a flexible (feasible) generalized least squares estimator on equation (6) where the assumptions are relaxed. However, there are no direct relationships with two-step estimation; a minor problem if the outset is equation (6).

Finally, note that consistency of the estimator(s) of β only requires n (the total number of observations) to grow large, whereas consistency of the estimators of α hinges on the number of groups, say g. To get consistent estimators of γ_j, it is clear that one would need the group sizes, say n_j, to grow.

3 Illustration

To illustrate our results results, we employ the various estimators in an analysis of the association between individual-level *fee-for-service expenditures* (FFSE) and co-morbidity of type 2 diabetes patients. We use a unique data set that allows us to combine information on the patients' morbidity with characteristics of the GP clinics providing treatment.

Danish GPs are self-employed professionals who are paid by regional governments (Olejaz et al., 2012). The current remuneration system for GPs, in which GPs are compensated through a combination of per capita fees (30%) and FFS (70%), does not differentiate the per capita component or the fees. However, the mixed GP remuneration system and other central parts of the primary care sector are currently undergoing a restructuring process (Pedersen et al., 2012; OECD Health Division, 2013). The success of these reforms and efforts to improve quality and efficiency will depend upon radically developing the data infrastructure underpinning primary care. One new element is that many Danish GPs have started to use the International Classification of Primary Care code (ICPC-2); see Schroll et al. (2008); Schroll (2009); WONCA (2005). The plan is that these data, which are routinely electronically collected by the Danish Quality Unit of Primary Care (DAK-E), will allow GPs to improve their quality of care. From 2013, the General Practitioners Organization (PLO) has agreed that all GPs will start diagnoses coding chronic patients such as Diabetes patients (Schroll et al., 2012). This development is in line with an international trend towards orienting resource allocation systems according to patients' overall health care needs (Starfield and Kinder, 2011). From an international perspective these new patient-level morbidity data combined with data on GP clinic activity and politically negotiated FFS tariffs offer a unique opportunity to explore how effectively the allocation of resources for FFS remuneration meets the health care needs. Therefore, in addition to the illustration of our theoretical results, our example provides a first attempt to explore the degree to which the proportion of FFSE variation is explained by patient morbidity and GP clinic characteristics for type 2 diabetes patients.

Our data set includes 6,706 type 2 diabetes patients who were registered and received services in 59 so-called *sentinel* GP clinics in 2010. These sentinel clinics are defined as those that coded more than 70% of their patients and are preferred for research and monitoring.[6] The dependent variable is patient-level FFSE defined as the sum of GP services weighted by politically negotiated service-specific fees. To measure the patients' morbidity we use simplified morbidity categories, termed Resource Utilization Bands (RUBs), based on the Adjusted Clinical Groups (ACG) case mix system developed by The Health Services Research & Development Center at The Johns Hopkins University (2009). The six mutually excluding RUBs are formed by combining the ACGs based on the patients' age, sex, and diagnoses codes. The ACG system software assigns the co-morbidity measures as listed in

[6]There might be a selection issue if the coding decision cannot be regarded as random. Explicitly testing for this would be an interesting pursuit when coding is no longer optional and data become available.

Table 1.

[Table 1 about here.]

Since higher morbidity could affect expenditures through an increased number of GP visits, we include the latter as an explicit control in the regressions. The GP-level variables we include are *clinic size*, the number of doctors in the clinics, the average *doctor's age* at the clinics, the *proportion of female doctors*, and the *number of diabetes patients per doctor*. Some descriptive statistics are given in Table 2, and the regression results from OLS and (feasible) GLS estimation of (6) and the two-stage estimators with different weighting schemes are given in Table 3.

[Table 2 about here.]

[Table 3 about here.]

It is clear that the different estimation strategies yield identical estimates of $\boldsymbol{\beta}$, the parameters associated with individual-level characteristics. The estimated parameters indicate that there are increasing expenditures associated with the degree of morbidity which becomes both statistically and practically significant as morbidity increases. Not surprisingly, the number of visits is also positively associated with expenditure.

Interestingly, none of the GP-level variables are significant in the regressions. However, one can confirm that the GLS estimates are identical to the second-stage WLS regression with $\widetilde{\omega}$ as weights[7], and that the OLS estimates are identical to the second-stage WLS regression with ω as weights. The second-stage OLS regression

[7]Of course, this requires that we use the same estimates of σ_u^2 and σ_v^2.

does not correspond to any of the other columns, although the practical difference is small in this case.

The group means ($\bar{\boldsymbol{x}}$) often have useful interpretations. In the current application the averaged RUB categories represents the proportions of diabetes patients the GPs have in each category. The averaged number of visits represent how many visits diabetes patients have on average in each GP. Only RUB2 and #visits are significant, but more interestingly: a (robust) Wald-test of $\boldsymbol{\pi} = \mathbf{0}$ yields a χ^2 statistic of 16.0 with a p-value of 0.0137 based on the GLS estimation (the conclusion based on the OLS estimation is very similar). The evidence illustrates a situation where we cannot simply combine equations (1) and (2), but should use either a two-stage approach including the group averages or a Mundlak-type estimation (since they are equivalent).

The variance component estimates are $\hat{\sigma}_v \approx 53.34$ and $\sigma_u \approx 129.97$. The ratio $\hat{\sigma}_v^2/(\hat{\sigma}_u^2 + \hat{\sigma}_v^2) \approx 0.14$ indicates that most of the unexplained variation in FFSE belongs to the individual level. This is interesting as almost all of our explanatory power also belongs to the individual level. $R^2 \approx 0.67$ (for both Mundlak-type estimations), which means that quite a large proportion of the variation can be explained for this patient type. If we omit the GP-level characteristics and include only individual characteristics and their aggregations (in terms of $\bar{\boldsymbol{x}}$), this drops to 0.66, a negligible decrease. Interpreting the variation decomposition based on the combined estimation has the benefit that we do not need to worry about the effect of estimation error in connection with $\hat{\gamma}_j$ and the second-stage regression.

The above results provide initial evidence that clinic-level information is not crucial for modeling FFSE, a result which is not obvious. However, the test rejecting $\boldsymbol{\pi} = 0$ suggests that clinic-level effects are correlated with morbidity and number of visits, and this should be controlled for. It also suggests that aggregated patient information as a measure of patient mix is important.

4 Conclusion and final remarks

Our theoretical results provide new insight to the random-intercept model and contribute to the field in several ways. First, it unifies several estimation procedures by introducing a single encompassing regression equation based on the Mundlak device. We show that its features carry over from the balanced panel data setup to the nested two-level setup. The equivalence results make it clear which assumptions need to be imposed as each provides certain insights. Secondly, certain weighting schemes for the second-level regression equation are highlighted and shown to be incorporated "automatically" when basing estimation on the new combined equation. Thirdly, a test for the dependence between individual characteristics and unobserved group effects becomes directly available as a bi-product of the new specification. Finally, we illustrate the results by analyzing how information about GP clinics and the morbidity of their diabetes patients explains FFS expenditure. Our data provide a unique opportunity to combine patient-level data with the GP clinics providing treatment. Our findings suggest that it is mainly information pertaining to the patients that explains FFS expenditure variation, at least for this category of patients. When more data become available this would require a more in-depth analysis to confirm these findings.

We did not consider potential endogeneity issues related to the unobserved individual-level component but have kept focus on issues related to the group-level component. The former is surely an important issue in many applications. Having formulated both levels of the model in a single encompassing regression equation makes many common econometric tools available for which applicability in the two-step approach is not obvious. Another topic interesting for future research is how our results extend to deeper hierarchical structures.

A Verification of the theoretical results

In the verification of the results it is helpful to recall a generalized Frisch-Waugh theorem; see Krishnakumar (2006). Consider the general partitioned regression:

$$\boldsymbol{y} = \boldsymbol{X}_1\boldsymbol{\beta}_1 + \boldsymbol{X}_2\boldsymbol{\beta}_2 + \boldsymbol{u}. \tag{8}$$

Then, for positive definite covariance matrix $\boldsymbol{\Omega}$,

$$\hat{\boldsymbol{\beta}}_{2,\text{GLS}} = (\boldsymbol{R}_2'\boldsymbol{\Omega}^{-1}\boldsymbol{R}_2)^{-1}\boldsymbol{R}_2'\boldsymbol{\Omega}^{-1}\boldsymbol{R}_1 \tag{9}$$

$$= (\boldsymbol{R}_2'\boldsymbol{\Omega}^{-1}\boldsymbol{R}_2)^{-1}\boldsymbol{R}_2'\boldsymbol{\Omega}^{-1}\boldsymbol{y}, \tag{10}$$

where

$$\boldsymbol{R}_1 = \boldsymbol{y} - \boldsymbol{X}_1(\boldsymbol{X}_1'\boldsymbol{\Omega}^{-1}\boldsymbol{X}_1)^{-1}\boldsymbol{X}_1'\boldsymbol{\Omega}^{-1}\boldsymbol{y} \tag{11}$$

$$\boldsymbol{R}_2 = \boldsymbol{X}_2 - \boldsymbol{X}_1(\boldsymbol{X}_1'\boldsymbol{\Omega}^{-1}\boldsymbol{X}_1)^{-1}\boldsymbol{X}_1'\boldsymbol{\Omega}^{-1}\boldsymbol{X}_2. \tag{12}$$

Also, if $\boldsymbol{X}_1'\Omega^{-1}\boldsymbol{X}_2 = \mathbf{0}$, then

$$\hat{\boldsymbol{\beta}}_{1,gls} = (\boldsymbol{X}_1'\Omega^{-1}\boldsymbol{X}_1)^{-1}\boldsymbol{X}_1'\Omega^{-1}\boldsymbol{y} \tag{13}$$

$$\hat{\boldsymbol{\beta}}_{2,gls} = (\boldsymbol{X}_2'\Omega^{-1}\boldsymbol{X}_2)^{-1}\boldsymbol{X}_2'\Omega^{-1}\boldsymbol{y}. \tag{14}$$

Note that the matrix Ω is the same in all of the regressions.

In what follows it is convenient to write the model in matrix notation as

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}\boldsymbol{\alpha} + \bar{\boldsymbol{X}}\boldsymbol{\pi} + \boldsymbol{v} + \boldsymbol{u}. \tag{15}$$

Let g be the number of groups, each with n_j members, $j = 1,\ldots,g$, and let $n = \sum n_j$. Define $\boldsymbol{P}$ to be a block-diagonal matrix with g blocks, each of size $n_j \times n_j$ with all entries being $1/n_j$, $j = 1,\ldots,g$. Also, let $\boldsymbol{Q} = \boldsymbol{I}_n - \boldsymbol{P}$; this is the well-known within transformation matrix. The *random effects* GLS variance-covariance matrix of the combined error term $\boldsymbol{v} + \boldsymbol{u}$ is then $\Omega = \boldsymbol{D}\boldsymbol{P} + \sigma_u^2\boldsymbol{Q}$, where $\boldsymbol{D}$ is an $n \times n$ diagonal matrix with values $n_j\sigma_v^2 + \sigma_u^2$ (note that the n_j term in the iith element, $i = 1,\ldots,n$, corresponds to the size of the group in which individual i is a member). It is easy to verify that $\Omega^{-1} = \boldsymbol{D}^{-1}\boldsymbol{P} + \sigma_u^{-2}\boldsymbol{Q}$, since $\Omega^{-1}\Omega = \boldsymbol{P} + \boldsymbol{Q} = \boldsymbol{I}$.

Now, noting that $\boldsymbol{X} = \boldsymbol{P}\boldsymbol{X} + \boldsymbol{Q}\boldsymbol{X}$ and that $\bar{\boldsymbol{X}} = \boldsymbol{P}\boldsymbol{X}$, we can rewrite (15) as

$$\boldsymbol{y} = \boldsymbol{Q}\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}\boldsymbol{\alpha} + \boldsymbol{P}\boldsymbol{X}(\boldsymbol{\pi} + \boldsymbol{\beta}) + \boldsymbol{v} + \boldsymbol{u} \tag{16}$$

$$= \boldsymbol{Q}\boldsymbol{X}\boldsymbol{\beta} + \widetilde{\boldsymbol{Z}}\boldsymbol{\delta} + \boldsymbol{v} + \boldsymbol{u}, \tag{17}$$

where

$$\widetilde{\boldsymbol{Z}} = [\boldsymbol{Z} \vdots \boldsymbol{P}\boldsymbol{X}] \tag{18}$$

$$\boldsymbol{\delta} = [\boldsymbol{\alpha}' \vdots (\boldsymbol{\pi} + \boldsymbol{\beta})']'. \tag{19}$$

To verify part (i), note that $\widetilde{\boldsymbol{Z}} = \boldsymbol{P}\widetilde{\boldsymbol{Z}}$, so

$$\boldsymbol{X}'\boldsymbol{Q}\Omega^{-1}\widetilde{\boldsymbol{Z}} = \boldsymbol{X}'\boldsymbol{Q}\Omega^{-1}\boldsymbol{P}\widetilde{\boldsymbol{Z}} \tag{20}$$

$$= \sigma_u^{-2}\boldsymbol{X}'\boldsymbol{Q}\boldsymbol{P}\widetilde{\boldsymbol{Z}} \tag{21}$$

$$= \boldsymbol{0} \tag{22}$$

since $\boldsymbol{Q}$ and $\boldsymbol{P}$ are orthogonal and $\boldsymbol{Q}$ is idempotent (as is $\boldsymbol{P}$). Therefore, from the generalized Frisch-Waugh theorem we have

$$\hat{\boldsymbol{\beta}}_{\text{GLS}} = (\boldsymbol{X}'\boldsymbol{Q}\Omega^{-1}\boldsymbol{Q}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Q}\Omega^{-1}\boldsymbol{y} \tag{23}$$

$$= (\boldsymbol{X}'\boldsymbol{Q}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Q}\boldsymbol{y} \tag{24}$$

$$= \hat{\boldsymbol{\beta}}_{\text{OLS}} = \hat{\boldsymbol{\beta}}_w. \tag{25}$$

The step from (23) to (24) can be realized using $\boldsymbol{Q}\boldsymbol{\Omega}^{-1} = \boldsymbol{Q}\boldsymbol{\Omega}^{-1}\boldsymbol{Q} = \sigma_u^{-2}\boldsymbol{Q}$.

To verify part (ii), we have from the classical Frisch-Waugh theorem that

$$\hat{\boldsymbol{\delta}}_{\text{OLS}} = (\widetilde{\boldsymbol{Z}}'\widetilde{\boldsymbol{Z}})^{-1}\widetilde{\boldsymbol{Z}}'\boldsymbol{y} \tag{26}$$

$$= (\widetilde{\boldsymbol{Z}}'\widetilde{\boldsymbol{Z}})^{-1}\widetilde{\boldsymbol{Z}}'(\boldsymbol{y} - \boldsymbol{Q}\boldsymbol{X}\hat{\boldsymbol{\beta}}_{\text{OLS}}) \tag{27}$$

$$= (\widetilde{\boldsymbol{Z}}'\widetilde{\boldsymbol{Z}})^{-1}[\boldsymbol{P}\widetilde{\boldsymbol{Z}}]'(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}_w + \boldsymbol{P}\boldsymbol{X}\hat{\boldsymbol{\beta}}_w) \tag{28}$$

$$= (\widetilde{\boldsymbol{Z}}'\widetilde{\boldsymbol{Z}})^{-1}\widetilde{\boldsymbol{Z}}'\hat{\boldsymbol{\gamma}} + (\widetilde{\boldsymbol{Z}}'\widetilde{\boldsymbol{Z}})^{-1}\widetilde{\boldsymbol{Z}}'\bar{\boldsymbol{X}}\hat{\boldsymbol{\beta}}_w \tag{29}$$

$$= (\hat{\boldsymbol{\alpha}}'_\omega, \hat{\boldsymbol{\pi}}'_\omega)' + (\boldsymbol{0}, \hat{\boldsymbol{\beta}}'_w)' \tag{30}$$

$$= \begin{pmatrix} \hat{\boldsymbol{\alpha}}_\omega \\ \hat{\boldsymbol{\pi}}_\omega + \hat{\boldsymbol{\beta}}_w \end{pmatrix}. \tag{31}$$

The last equality holds since we end up with a regression of averaged data with $\hat{\boldsymbol{\gamma}}$ as dependent variable. All observations within each group are identical so this is the same as a weighted regression based on the unique averaged observations with weights ω_j.

Part (iii) is verified similarly:

$$\hat{\boldsymbol{\delta}}_{\text{GLS}} = (\widetilde{\boldsymbol{Z}}'\boldsymbol{\Omega}^{-1}\widetilde{\boldsymbol{Z}})^{-1}\widetilde{\boldsymbol{Z}}'\boldsymbol{\Omega}^{-1}\boldsymbol{y} \tag{32}$$

$$= (\widetilde{\boldsymbol{Z}}'\boldsymbol{\Omega}^{-1}\widetilde{\boldsymbol{Z}})^{-1}\widetilde{\boldsymbol{Z}}'\boldsymbol{\Omega}^{-1}(\boldsymbol{y} - \boldsymbol{Q}\boldsymbol{X}\hat{\boldsymbol{\beta}}_{\text{GLS}}) \tag{33}$$

$$= (\widetilde{\boldsymbol{Z}}'\boldsymbol{\Omega}^{-1}\widetilde{\boldsymbol{Z}})^{-1}\widetilde{\boldsymbol{Z}}'\boldsymbol{\Omega}^{-1}[\boldsymbol{P} + \boldsymbol{Q}](\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}_w + \boldsymbol{P}\boldsymbol{X}\hat{\boldsymbol{\beta}}_w) \tag{34}$$

$$= (\widetilde{\boldsymbol{Z}}'\boldsymbol{\Omega}^{-1}\widetilde{\boldsymbol{Z}})^{-1}\widetilde{\boldsymbol{Z}}'\boldsymbol{\Omega}^{-1}\boldsymbol{P}(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}_w + \boldsymbol{P}\boldsymbol{X}\hat{\boldsymbol{\beta}}_w) \tag{35}$$

$$= (\widetilde{\boldsymbol{Z}}'\boldsymbol{\Omega}^{-1}\widetilde{\boldsymbol{Z}})^{-1}\widetilde{\boldsymbol{Z}}'\boldsymbol{\Omega}^{-1}\hat{\boldsymbol{\gamma}} + (\widetilde{\boldsymbol{Z}}'\boldsymbol{\Omega}^{-1}\widetilde{\boldsymbol{Z}})^{-1}\widetilde{\boldsymbol{Z}}'\boldsymbol{\Omega}^{-1}\bar{\boldsymbol{X}}\hat{\boldsymbol{\beta}}_w \tag{36}$$

$$= (\hat{\boldsymbol{\alpha}}'_{\widetilde{\omega}}, \hat{\boldsymbol{\pi}}'_{\widetilde{\omega}})' + (\boldsymbol{0}, \hat{\boldsymbol{\beta}}'_w)' \tag{37}$$

$$= \begin{pmatrix} \hat{\boldsymbol{\alpha}}_{\widetilde{\omega}} \\ \hat{\boldsymbol{\pi}}_{\widetilde{\omega}} + \hat{\boldsymbol{\beta}}_w \end{pmatrix}. \tag{38}$$

The last equality follows from the definition of $\boldsymbol{D}$, implying that the diagonal elements of its inverse are $1/[n_j\sigma_v^2 + \sigma_u^2]$, which combined with the n_j replications of each unique observation makes the last expression correspond to a second-stage regression of unique averaged observations with $\hat{\gamma}$ as dependent variable and weights $n_j/[n_j\sigma_v^2 + \sigma_u^2] = \sigma_v^2/[1 + (\sigma_u^2/\sigma_u^2)/n_j] \propto \tilde{\omega}_j$.

Part (iv) is obvious when considering equations (6) and (7) and the usual requirements for unbiasedness and consistency of least squares estimators.

B Estimation with quasi-demeaned data

Here, we review a practical approach to feasible GLS estimation of equation (6) where data is *quasi-demeaned*, an approach which is also valid in the current setting. GLS estimation amounts to pre-multiplying data columns with $\sigma_u\boldsymbol{\Omega}^{-1/2}$ and applying OLS on the transformed data. It is straightforward to verify that $\boldsymbol{\Omega}^{-1/2} = \boldsymbol{D}^{-1/2}\boldsymbol{P} + \sigma_u^{-1}\boldsymbol{Q}$, so that $\sigma_u\boldsymbol{\Omega}^{-1/2} = \sigma_u\boldsymbol{D}^{-1/2}\boldsymbol{P} + \boldsymbol{Q}$. Define

$$\theta_j = 1 - \sqrt{\frac{\sigma_u^2}{n_j\sigma_v^2 + \sigma_u^2}} \tag{39}$$

$$= 1 - \sqrt{\frac{1}{n_j(\sigma_v^2/\sigma_u^2) + 1}} \tag{40}$$

and note that elements of the diagonal matrix $\sigma_u\boldsymbol{D}^{-1/2}$ are of the form $1 - \theta_j$.

Now, consider e.g. the transformation of the dependent variable:

$$\sigma_u \Omega^{-1/2} \boldsymbol{y} = \sigma_u \boldsymbol{D}^{-1/2} \boldsymbol{P} \boldsymbol{y} + \boldsymbol{Q} \boldsymbol{y} \tag{41}$$

$$= \sigma_u \boldsymbol{D}^{-1/2} \boldsymbol{P} \boldsymbol{y} - \boldsymbol{P} \boldsymbol{y} + \boldsymbol{y} \tag{42}$$

$$= \sigma_u \boldsymbol{D}^{-1/2} \bar{\boldsymbol{y}} - \bar{\boldsymbol{y}} + \boldsymbol{y} \tag{43}$$

$$= \boldsymbol{y} - (1 - \sigma_u \boldsymbol{D}^{-1/2}) \bar{\boldsymbol{y}}, \tag{44}$$

where the bar denotes a vector with group averages. From this one obtains that each transformed observation is of the form $y_i - \theta_j \bar{y}_j$. Similar transformation applies to the independent variables, too. The transformation subtracts from each observation a fraction of the group mean, and this is called *quasi-demeaning*. It is interesting to note that $n_j \to \infty$ implies $\theta_j \to 1$, so as all group sizes grow large, GLS estimation tends to the within estimator.

The variance components σ_u^2 and σ_v^2 are unknown, and hence the θ_js are unknown, but these quantities can be estimated leading to the feasible estimator. Let $a_{ij} = v_j + u_i$, and note that $\mathrm{E}[a_{ij} a_{lj}] = \sigma_v^2$ for $i \neq l$. Using the non-redundant observations, this suggests the estimator

$$\hat{\sigma}_v^2 = \frac{1}{m-p} \sum_{j=1}^{g} \sum_{i=1}^{n_j - 1} \sum_{l=i+1}^{n_j} \hat{a}_{ij} \hat{a}_{lj}, \quad \text{where} \tag{45}$$

$$m = \sum_{j=1}^{g} n_j (n_j - 1)/2, \tag{46}$$

$\hat{a}_{ij}$ are the residuals from an OLS estimation, and p is the total number of parameters in α, β, and π. Note that index i is used differently here where it ranges from 1 to n_j for each j, rather than from 1 to n. This is convenient for writing the sum of the relevant cross-products.

Now, since $\sigma_a^2 = \text{var}(a_{ij}) = \sigma_u^2 + \sigma_v^2$, we can estimate of σ_u^2 as

$$\hat{\sigma}_u^2 = \hat{\sigma}_a^2 - \hat{\sigma}_v^2, \quad \text{where} \tag{47}$$

$$\hat{\sigma}_a^2 = \frac{1}{n-p} \sum_{j=1}^{g} \sum_{i=1}^{n_j} \hat{a}_{ij}^2. \tag{48}$$

There are other approaches to estimating the variance components. One alternative uses the so-called *between* estimator. This approach will not work here, since it is based on group averages of all variables inducing singularity since $\bar{\boldsymbol{X}}$ will then be included twice in the design matrix. Finally, under the assumption imposed on the variance structure, the variance-covariance matrix for inference can be estimated as

$$\text{var}(\hat{\boldsymbol{\xi}}) = n(\boldsymbol{H}'\hat{\Omega}^{-1}\boldsymbol{H})^{-1}, \quad \text{where} \tag{49}$$

$$\hat{\boldsymbol{\xi}} = (\hat{\boldsymbol{\beta}}', \hat{\boldsymbol{\alpha}}', \hat{\boldsymbol{\pi}}')' \quad \text{and} \tag{50}$$

$$\boldsymbol{H} = [\boldsymbol{X} \vdots \boldsymbol{Z} \vdots \bar{\boldsymbol{X}}]. \tag{51}$$

Alternatively, one can use robust versions if one suspects the imposed variance structure to be incorrect.

References

Baltagi, B. H. (2006). An alternative derivation of Mundlak's fixed effects results using system estimation. *Econometric Theory 22*(6), 1191.

Blundell, R. and F. Windmeijer (1997). Cluster effects and simultaneity in multilevel models. *Health Economics 6*(4), 439–443.

Carey, K. (2000). A multilevel modelling approach to analysis of patient costs under managed care. *Health Economics 9*(5), 435–446.

Fletcher, J. M. (2010). Social interactions and smoking: Evidence using multiple student cohorts, instrumental variables, and school fixed effects. *Health Economics 19*(4), 466–484.

Gurka, M. J., L. J. Edwards, and K. E. Muller (2011). Avoiding bias in mixed model inference for fixed effects. *Statistics in Medicine 30*(22), 2696–2707.

Hausman, J. (1978). Specification tests in econometrics. *Econometrica: Journal of the Econometric Society 46*(6), 1251–1271.

Krishnakumar, J. (2006). *Time Invariant Variables and Panel Data Models: A Generalised Frisch Waugh Theorem and its Implications*, Volume 274. Emerald Group Publishing Limited.

Laudicella, M., K. R. Olsen, and A. Street (2010). Examining cost variation across hospital departments–a two-stage multi-level approach using patient-level data. *Social Science & Medicine 71*(10), 1872–1881.

Lewis, J. B. and D. A. Linzer (2005). Estimating regression models in which the dependent variable is based on estimates. *Political Analysis 13*(4), 345–364.

Mundlak, Y. (1978). On the pooling of time series and cross section data. *Econometrica: journal of the Econometric Society 46*(1), 69–85.

Olejaz, M., A. Juul Nielsen, A. Rudkjøbing, B. H. Okkels, A. Kransik, and C. Hernández-Ouevedo (2012). Denmark: Health system review. *Health Systems in Transitition 14*(2), 1–192.

Pedersen, K., J. Andersen, and J. Søndergaard (2012). General Practice and Primary Health Care in Denmark. *Journal of the American Board of Family Medicine 25*, 34–38.

Rice, N. and A. Jones (1997). Multilevel models and health economics. *Health Economics 6*, 561–75.

Schroll, H. (2009). Data collection perspectives from patient care in general practice. *Ugeskr Laeger 171*(20), 1681–1684.

Schroll, H., B. Christensen, J. S. Andersen, and J. Søndergaard (2008). Danish General Medicine Database–future tool! The Danish Society of General Medicine. *Ugeskr Laeger 170*(12), 1013.

Schroll, H., R. D. Christensen, J. L. Thomsen, M. Andersen, S. Friborg, and J. Søndergaard (2012). The Danish model for improvement of diabetes care in general practice: impact of automated collection and feedback of patient data. *International Journal of Family Medicine 2012*, 1–5.

Scribner, R. A., K. P. Theall, N. R. Simonsen, K. E. Mason, and Q. Yu (2009). Misspecification of the effect of race in fixed effects models of health inequalities. *Social Science & Medicine 69*(11), 1584–1591.

Starfield, B. and K. Kinder (2011). Multimorbidity and its measurement. *Health Policy 103*(1), 3–8.

OECD Health Division (2013). OECD reviews of health care quality: Denmark - executive summary; assessment and recommendations. OECD *Better policies for better lives*, 1–34.

WONCA (2005). *International Classification of Primary Care. ICPC-2-R.* (Second ed.). Oxford University Press, New York.

The Health Services Research & Development Center at The Johns Hopkins University (2009). The Johns Hopkins University. The Johns Hopkins ACG® system, Technical Reference Guide Version 9.0. Bloomberg School of Public Health.

RUB0:	non-users
RUB1:	healthy users
RUB2:	low morbidity
RUB3:	moderate morbidity
RUB4:	high morbidity
RUB5:	very high morbidity

Table 1: Resource Utilization Band categories.

	Variable	Mean	5%	Median	75%	95%	99%
Individual level	FFSE (€)	398·48	96·67	353·26	500·89	841·03	1,200·34
	RUB0	0·133					
	RUB1	0·081					
	RUB2	0·264					
	RUB3	0·478					
	RUB4	0·040					
	RUB5	0·004					
	#visits	7·844	1	6	10	20	31
GP level	Clinic size	3·248	1	3	4	7	8
	Average GP age	53·230	42·00	53·33	56·67	65·50	74·50
	Prop. of female GP	0·514	0·00	0·50	0·67	1·00	1·00
	#diab. patients/GP	56·989	24·00	52·33	73·00	101·00	114·00

Table 2: Descriptive statistics. The sample includes 59 GPs with a total of 6,702 type 2 diabetes patients.

		Combined estimation		2-step estimation		
		OLS	GLS	WLS, ω	WLS, $\widetilde{\omega}$	OLS
Individual characteristics, x and β	RUB1	1·138	1·138	1·138	1·138	1·138
		(6·999)	(6·999)	(6·994)	(6·994)	(6·994)
	RUB2	11·22*	11·22*	11·22*	11·22*	11·22*
		(6·236)	(6·236)	(6·232)	(6·232)	(6·232)
	RUB3	46·642***	46·642***	46·642***	46·642***	46·642***
		(8·881)	(8·881)	(8·874)	(8·874)	(8·874)
	RUB4	98·886***	98·886***	98·886***	98·886***	98·886***
		(15·381)	(15·381)	(15·37)	(15·37)	(15·37)
	RUB5	221·636***	221·636***	221·636***	221·636***	221·636***
		(61·197)	(61·197)	(61·151)	(61·151)	(61·151)
	#visits	26·019***	26·019***	26·019***	26·019***	26·019***
		(0·929)	(0·929)	(0·929)	(0·929)	(0·929)
GP characteristics, z and α	Intercept	85·759	112·965	85·759	112·964	107·206
		(113·47)	(135·46)	(124·582)	(148·726)	(153·084)
	Clinic size	10·013**	8·75*	10·013**	8·75	8·862
		(4·228)	(5·279)	(4·642)	(5·796)	(5·916)
	Average GP age	0·111	−0·327	0·111	−0·327	−0·406
		(1·268)	(1·197)	(1·393)	(1·314)	(1·305)
	Prop. of women GP	26·769	27·314	26·769	27·314	28·292
		(22·035)	(20·429)	(24·192)	(22·43)	(22·469)
	#Diab. patients./GP	0·264	0·512*	0·264	0·512*	0·539*
		(0·244)	(0·276)	(0·268)	(0·303)	(0·312)
Mundlak device, $\bar{x}$ and π	$\overline{\text{RUB1}}$	328·813	236·223	328·813	236·223	243·602
		(217·547)	(249·676)	(239·186)	(274·365)	(281·101)
	$\overline{\text{RUB2}}$	−423·017**	−324·295	−423·017**	−324·295	−317·241
		(178·802)	(197·969)	(196·225)	(217·213)	(224·141)
	$\overline{\text{RUB3}}$	30·914	37·089	30·914	37·089	51·779
		(95·45)	(125·397)	(104·642)	(136·377)	(142·591)
	$\overline{\text{RUB4}}$	145·002	−2·563	145·002	−2·563	−10·897
		(240·288)	(229·259)	(263·961)	(252·047)	(251·545)
	$\overline{\text{RUB5}}$	−1531·374	−1731·358	−1531·374	−1731·358	−1780·114
		(1022·831)	(1091·922)	(1146·685)	(1222·336)	(1237·038)
	$\overline{\text{\#visits}}$	10·422***	7·039**	10·422***	7·039**	6·87*
		(2·166)	(2·937)	(2·807)	(3·483)	(3·537)

Table 3: Regression results from the combined Mundlak-type equation using ordinary least squares (OLS) and feasible generalized least squares (GLS) along with the 2-stage procedures where the first stage is based on the *within* estimator and the second stage is estimated by weighted least squares (WLS) and OLS. The dependent variable is *fee-for-service expenditures*. Standard errors are robust to clustering and heteroskedasticity. Significance codes are $^{***}: p < 0.01$, $^{**}: p < 0.05$, $^{*}: p < 0.1$.