pdynmc: A Package for Estimating Linear Dynamic Panel Data Models Based on Nonlinear Moment Conditions

by Markus Fritsch, Andrew Adrian Yu Pua and Joachim Schnurbus

Abstract This paper introduces **pdynmc**, an R package that provides users sufficient flexibility and precise control over the estimation and inference in linear dynamic panel data models. The package primarily allows for the inclusion of nonlinear moment conditions and the use of iterated GMM; additionally, visualizations for data structure and estimation results are provided. The current implementation reflects recent developments in literature, uses sensible argument defaults, and aligns commercial and noncommercial estimation commands. Since the understanding of the model assumptions is vital for setting up plausible estimation routines, we provide a broad introduction of linear dynamic panel data models directed towards practitioners before concisely describing the functionality available in **pdynmc** regarding instrument type, covariate type, estimation methodology, and general configuration. We then demonstrate the functionality by revisiting the popular firm-level dataset of Arellano and Bond (1991).

Introduction

This paper introduces the contributed package **pdynmc** (Fritsch et al., 2020) – a unified framework for estimating linear dynamic panel data models based on linear and nonlinear moment conditions (Ahn and Schmidt, 1995). Our implementation of the commands in **pdynmc** allows the user to exert precise control over the available functionality, reflects recent developments in literature, uses sensible argument defaults, aligns commercial and noncommercial estimation commands, and provides visualizations of data structure and estimation results. Additionally, this paper provides a concise introduction into linear dynamic panel data models directed towards the practitioner, describes the functionality available in **pdynmc**, and walks the reader through estimation of linear dynamic panel data models by replicating the analysis in Arellano and Bond (1991).

Practitioners have a variety of recent packages that enable linear dynamic panel data modeling meant for a fixed number of time periods. In particular, contributed R packages such as **OrthoPanels** (Pickup et al., 2017), **plm** (Croissant and Millo, 2019), and **panelvar** (Sigmund and Ferstl, 2019) have considerably enlarged the set of noncommercial routines available. All of these packages implement some default routines for estimating common parameters in linear dynamic panel data models. **OrthoPanels** implements a likelihood-based orthogonal reparameterization procedure for first-order autoregressive linear panel data models with strictly exogenous covariates. **plm** implements one-step and two-step GMM-based procedures for pth-order autoregressive linear panel data models. **panelvar** implements iterated (or "m-step", compare Sigmund and Ferstl, 2019) GMM procedures for pth-order vector autoregressive linear panel data models. For the latter two packages, linear moment conditions are used to identify common parameters.

Additional functionality of our contributed package **pdynmc** includes nonlinear moment conditions which are generally not available across standard GMM estimation routines. To the best of our knowledge, there is currently only the xtdpdgmm-implementation provided by (Kripfganz, 2019) for the commercial statistical software Stata (StataCorp, 2015). The current implementation in Stata restricts accessibility to the routine as it requires a recent Stata version (version 13 or higher). Another key estimation option provided by **pdynmc** is iterated GMM; Hansen and Lee (2021) recently outlined the merits of the technique and developed the theory under potential misspecification of moment conditions. The availability of iterated GMM for dynamic panels may help to apply the results found in, for example, Hwang and Sun (2018). Visualizations of the estimation results of iterated GMM are also included.

The structure of the paper is as follows. Section Framework and methodology briefly sketches the linear dynamic panel data model, states underlying assumptions frequently used in literature, and describes moment conditions arising from different sets of model assumptions. Section R implementation covers details on the arguments of the model fitting function pdynmc and connects the description with the estimation methodology. Section Empirical example illustrates the estimation of linear dynamic panel data models with **pdynmc** for the data set of Arellano and Bond (1991), while Section Conclusion concludes.

Framework and methodology

Linear dynamic panel data models account for dynamics and unobserved individual-specific heterogeneity. Due to the presence of lagged dependent variables, applying ordinary least squares including individual-specific dummy variables is inconsistent (see, e.g., Hsiao, 2014). A suitable alternative for obtaining parameter estimates of linear dynamic panel data models is deriving moment conditions (or population orthogonality conditions) from the model assumptions. The moment conditions may be linear (Anderson and Hsiao, 1982; Holtz-Eakin et al., 1988; Arellano and Bover, 1995) or nonlinear (Ahn and Schmidt, 1995) in parameters and determine the natural instruments available for estimation. Usually, the number of moment conditions exceeds the number of parameters and moment conditions need to be aggregated appropriately. This can be achieved by the generalized method of moments (GMM), where weighted linear combinations of moment conditions are employed to obtain parameter estimates.

Theoretical results and evidence from Monte Carlo simulations in literature suggest that incorporating nonlinear (quadratic) moment conditions proposed by Ahn and Schmidt (1995) is valuable for identification: For example, when the lag parameter exhibits high persistence, linear moment conditions fail to identify the model parameters, while quadratic moment conditions can still provide identification (Bun and Kleibergen, 2021; Bun and Sarafidis, 2015; Gørgens et al., 2019; Pua et al., 2019a,b). Note that the quadratic moment conditions are immediate by-products of imposing standard assumptions, which are the basis of the Arellano and Bond (1991) estimator – the most popular default routine in dynamic panel data estimation.

Since the moment conditions employed in GMM estimation of linear dynamic panel data models are derived from model assumptions, a basic understanding of these assumptions is vital for setting up a plausible estimation routine. We briefly review the assumptions implied when using particular moment conditions in estimation below and add to the exposition in the **plm** vignette (Croissant and Millo, 2019), where the function pgmm is used to estimate linear dynamic panel data models. For further reading on the methodology, we suggest Fritsch (2019).

Linear dynamic panel data model

For a given data set with cross section dimension n and time series dimension T, consider a linear dynamic panel data model of the form:

$$y_{i,t} = \alpha y_{i,t-1} + \beta x_{i,t} + u_{i,t}, \quad i = 1, \dots, n; \ t = 2, \dots, T, \tag{1}$$
$$u_{i,t} = \eta_i + \varepsilon_{i,t}. \tag{2}$$

Variables $y_{i,t}$ and $y_{i,t-1}$ denote the dependent variable and its lag, α is the lag parameter, and $x_{i,t}$ is a single covariate with corresponding slope coefficient β. The second equation separates the composite error term $u_{i,t}$ into an unobserved individual-specific effect η_i and an idiosyncratic remainder component $\varepsilon_{i,t}$.

Combining Equations (1) and (2) yields the single equation form

$$y_{i,t} = \alpha y_{i,t-1} + \beta x_{i,t} + \eta_i + \varepsilon_{i,t}. \tag{3}$$

We only include one lag of the dependent variable, one covariate, and omit unobserved time-specific effects in this section for simplicity of exposition and notational convenience. Extending the representation is straightforward and **pdynmc** can also accommodate AR(p) models and time effects. The initial time period is denoted by $t = 1$.

The unobserved individual-specific effects η_i may be eliminated from Equation (3) by taking first differences:

$$\Delta y_{i,t} = \alpha \Delta y_{i,t-1} + \beta \Delta x_{i,t} + \Delta \varepsilon_{i,t}. \tag{4}$$

Since the first difference of the lagged dependent variable $\Delta y_{i,t-1} = y_{i,t-1} - y_{i,t-2}$ and the first difference of the idiosyncratic remainder component $\Delta \varepsilon_{i,t} = \varepsilon_{i,t} - \varepsilon_{i,t-1}$ are not orthogonal, ordinary least squares estimation of Equation (4) is inconsistent. Linear dynamic panel data models are usually estimated by GMM, where corresponding moment conditions are derived from model assumptions.

Standard assumptions and associated moment conditions

Researchers have focused on the following standard assumptions, henceforth StA, (see Ahn and Schmidt, 1995):

$$\text{The data are independently distributed across } i, \qquad (5)$$

$$E(\eta_i) = 0, \quad i = 1, ..., n,$$

$$E(\varepsilon_{i,t}) = 0, \quad i = 1, ..., n, \ t = 2, ..., T,$$

$$E(\varepsilon_{i,t} \cdot \eta_i) = 0, \quad i = 1, ..., n, \ t = 2, ..., T,$$

$$E(\varepsilon_{i,t} \cdot \varepsilon_{i,s}) = 0, \quad i = 1, ..., n, \ t \neq s,$$

$$E(y_{i,1} \cdot \varepsilon_{i,t}) = 0, \quad i = 1, ..., n, \ t = 2, ..., T,$$

$$n \to \infty, \text{ while } T \text{ is fixed, such that } \frac{T}{n} \to 0.$$

We assume that StA hold for the rest of this paper.

Under StA, Holtz-Eakin et al. (1988) (henceforth **HNR**) propose the moment conditions

$$E(y_{i,s} \cdot \Delta u_{i,t}) = 0, \qquad t = 3, \dots, T; \ s = 1, \dots, t-2. \qquad (6)$$

Equation (6) provides $0.5(T-1)(T-2)$ moment conditions. Equivalent moment conditions can be derived from the covariate $x_{i,t}$ – depending on its correlation with the idiosyncratic remainder component $\varepsilon_{i,t}$. Endogenous, predetermined, and (strictly) exogenous covariates provide the following linear moment conditions, respectively:

$$E(x_{i,s} \cdot \Delta u_{i,t}) = 0, \qquad t = 3, \dots, T, \qquad \text{where} \qquad (7)$$

$$s = 1, \dots, t-2, \quad \text{for} \quad \text{endogenous } x,$$

$$s = 1, \dots, t-1, \quad \text{for} \quad \text{predetermined } x,$$

$$s = 1, \dots, T, \quad \text{for} \quad \text{strictly exogenous } x.$$

After solving Equation (1) for $u_{i,t}$ and inserting this as $\Delta u_{i,t} = u_{i,t} - u_{i,t-1}$, it is apparent that the HNR moment conditions are linear in parameters (α and β). In literature, the HNR moment conditions also appear as "moment conditions with instruments in levels" (w.r.t. $y_{i,s}, x_{i,s}$) and "moment conditions from equations in differences" (w.r.t. $\Delta u_{i,t}$).

Ahn and Schmidt (1995) (henceforth **AS**) have shown that under StA the following $T - 3$ additional moment conditions hold:

$$E(u_{i,T} \cdot \Delta u_{i,t-1}) = 0, \qquad t = 4, \dots, T. \qquad (8)$$

These moment conditions are nonlinear in parameters (quadratic in α and β).

Extended assumptions and associated moment conditions

Another set of moment conditions, beyond those implied by StA, that is popular in theoretical and applied research is derived from the assumption

$$E(\Delta y_{i,t} \cdot \eta_i) = 0, \quad i = 1, \dots, n. \qquad (9)$$

This expression requires that the dependent variable and the unobserved individual-specific effects are constantly correlated over time for each individual and has thus been called "constant correlated effects" (Bun and Sarafidis, 2015). This assumption is also called "effect stationarity" (Kiviet, 2007) or "mean stationarity" (Arellano, 2003).

From this assumption, Arellano and Bover (1995) derive $T - 2$ additional moment conditions (henceforth **ABov**):

$$E(\Delta y_{i,t-1} \cdot u_{i,t}) = 0, \quad t = 3, \dots, T. \qquad (10)$$

By rewriting these moment conditions, it can be shown that the ABov moment conditions encompass the nonlinear AS moment conditions (for a derivation see Fritsch, 2019).

Depending on the assumptions about $x_{i,t}$, additional ABov moment conditions can be derived:

$$E(\Delta x_{i,v} \cdot u_{i,t}) = 0, \qquad \text{where}$$
$$v = t - 1; \ t = 3, \ldots T, \quad \text{for} \quad x \text{ endogenous,}$$
$$v = t; \ t = 2, \ldots, T, \quad \text{for} \quad x \text{ strictly exogenous} \quad \text{or} \quad x \text{ predetermined.}$$

Deviations from the assumption are required to be unsystematic over both, the cross section and the time series dimension (see Section 6.5 in Arellano, 2003, which also provides an empirically relevant example). Employing the constant correlated effects assumption implicitly constrains the relationship between $\Delta x_{i,t}$ and η_i (see Blundell et al., 2001). If the statistician is not willing to impose this restriction, nonlinear AS moment conditions can be used instead.

R implementation

Similar to function pgmm in the package **plm**, **pdynmc** provides one-step and two-step closed form GMM estimators and standard specification testing such as overidentifying restrictions tests, serial correlation tests, and Wald tests. These features are shared by other packages implemented in Gauss, Ox (Doornik et al., 2012), R, and Stata. We provide options to match results from other statistical software estimation routines. In contrast to **OrthoPanels** and **plm**, **pdynmc** does not include a formula interface to allow the user to exert full control over all functionality.

GMM estimation, inference, and testing

We provide one-step, two-step, and iterated estimation for the coefficients. The weighting matrix of the moment conditions plays a prominent role in estimation (Arellano and Bond, 1991; Blundell et al., 2001; Kripfganz, 2019). An optimal weighting matrix is proportional to the inverse of the covariance matrix of the moment conditions (see, e.g., Arellano, 2003). The default weighting matrix used in **pdynmc** is based on the proposal of Arellano and Bond (1991). For details on available alternatives, see the documentation of **pdynmc** and the corresponding package vignette (Fritsch et al., 2020).

Details on the computation of asymptotic one- and two-step standard errors can be found in Doornik et al. (2012). As asymptotic two-step GMM standard errors for the estimated coefficients exhibit a downward bias in small samples, they can be substantially lower than one-step GMM standard errors (see, e.g., Arellano and Bond, 1991). Windmeijer (2005) relates the bias to the dependence of the two-step weighting matrix on one-step parameter estimates and proposes an analytic correction of two-step standard errors. Robust and non-robust versions of the standard errors are available.

Coefficient estimates and standard errors from one- and two-step GMM estimation can be misleading. An example of high practical relevance is, when the estimated model is a reasonable approximation instead of the true functional relationship (Hansen and Lee, 2021; Hwang et al., 2021): This may render some of the moment conditions invalid. Hansen and Lee (2021) highlight the arbitrariness of the initial weighting matrix and note that iterated GMM provides a remedy. Across iterations, the weighting matrix is updated based on the residuals of the previous estimation step (for more details, see Hansen and Lee, 2021, p. 4–6). Iterated GMM is used as a default in **pdynmc**.

We implement the following tests:

- The serial correlation test of Arellano (2003).
- The overidentifying restrictions test of Hansen (1982), called "J-test".
- A Wald test of joint significance of (i) coefficients of lagged-dependent variable(s) and covariates; (ii) time dummy coefficients; (iii) both, (i) and (ii).

When nonlinear moment conditions are used in GMM estimation, nonlinear optimization techniques are required to obtain coefficient estimates. By default, GMM estimation by **pdynmc** is based on numerical optimization. For the optimization procedure, we rely on R-package **optimx** (Nash and Varadhan, 2011; Nash, 2014). All optimization routines implemented in **optimx** are available in **pdynmc**. Based on our experience, we recommend using the Variable Metric method (Fletcher, 1970; Nash, 1990, 2020) in estimation of linear dynamic panel data models. The technique is labeled BFGS in **optimx** and serves as default procedure in **pdynmc**.

Function arguments explained

In package **pdynmc**, the eponymous function is intended for model fitting. The most important function arguments for applied work are summarized in Table 1.

Argument	Description
`dat`	Data set with rows (i.e., observations) and columns (i.e., variables).
`varname.i`	Cross-section identifier (column name).
`varname.t`	Time-series identifier (column name).
`use.mc.diff`	Use moment conditions from Equation (6) (i.e., equations in differences, instruments in levels).
`use.mc.lev`	Use moment conditions from Equation (10) (i.e., equations in levels, instruments in differences).
`use.mc.nonlin`	Use nonlinear (quadratic) moment conditions from Equation (8).
`use.mc.nonlinAS`	If turned to `FALSE`, the nonlinear moment conditions are used in a modified version (that is also valid under StA), where T in Equation (8) is replaced by t.
`include.y`	Derive instruments from lags of dependent variable.
`varname.y`	Name of dependent variable in data set.
`lagTerms.y`	Number of lags of dependent variable.
`maxLags.y`	Maximum number of lags of dependent variable from which to derive instruments.
`include.x`	Derive instruments from covariates.
`varname.reg.end`	Name(s) of covariate(s) to be treated as endogenous (replace suffix end by `pre` for predetermined; by `ex` for exogenous covariates).
`lagTerms.reg.end`	Number of lags of endogenous covariate(s) (also for `pre` or `ex`).
`maxLags.reg.end`	Maximum number of lags of endogenous covariate(s) used to derive instruments (also for `pre` or `ex`).
`fur.con`	Include further control variables (i.e., covariates that are not used for deriving instruments).
`fur.con.diff`	Logical variable indicating whether to include further control variables in equations in differences.
`fur.con.lev`	Logical variable indicating whether to include further control variables in equations in levels.
`varname.reg.fur`	Name(s) of covariate(s) in data set to be treated as further controls.
`lagTerms.reg.fur`	Number of lags of further controls.
`include.dum`	Include time dummies. Note: Can be constructed from multiple variables.
`dum.diff`	Include time dummies in equations in first differences.
`dum.lev`	Include time dummies in equations in levels.
`varname.dum`	Variable name(s) for creating time dummies (can be different from `varname.t`).
`w.mat`	Type of weighting matrix to be used, `iid.err` (as proposed by Arellano and Bond (1991)), `identity`, or `zero.cov`.
`std.err`	Type of standard errors to be used, either bias-corrected (`corrected`) according to Windmeijer (2005) or not (`unadjusted`).
`estimation`	Type of estimation, `onestep`, `twostep`, or `iterative`.

Table 1: Most-used arguments of function pdynmc.

Besides the arguments described in Table 1, various further configuration options exist for function pdynmc. The function allows for inclusion of further covariates which are only used as instruments (i.e., covariates from which moment conditions are derived, but for which no parameters are estimated; compare arguments `include.x.instr` and `varname.reg.instr`) as well as the opposite, covariates which are instrumented (i.e., covariates for which parameters are estimated, but from which no moment conditions are derived; compare arguments `include.x.toInstr` and `varname.reg.toInstr`).

Further, thresholds for collinearity checks can be adjusted via `col_tol`. The total number of instruments above which a generalized inverse is used to invert the weighting matrix can be specified by `inst.thresh`.

When only linear moment conditions are used, a closed-form solution exists for the estimator and nonlinear optimization can be turned off (by `opt.meth = "none"`). Package **optimx** is employed for nonlinear optimization and argument `hessian` controls, whether the Hessian matrix is approximated in estimation; all other control arguments for **optimx** can be provided in a list via `optCtrl`. Starting values for initializing nonlinear optimization are drawn from the uniform distribution in the interval [-1, 1] via `start.val.lo = -1` and `start.val.hi = 1`; a seed (`seed.input = 42`) ensures reproducibility (limits and seed can be adjusted). Alternatively, specific starting values can be provided via argument `start.val` (note that `custom.start.val` has to be set to `TRUE`).

For iterated estimation, termination criteria can be set via `max.iter` (maximum number of iterations) and `iter.tol` (tolerance for determining convergence).

The Stata packages xtabond2 (Roodman, 2018) and xtdpdgmm have a somewhat different usage and weighting of moment conditions. If HNR and ABov moment conditions are available for estimation, some of the ABov moment conditions are redundant (see Fritsch, 2019, for a derivation). While the Stata routines fully expand the linear ABov moment conditions when setting up the instrument matrix (including the redundant moment conditions), **pdynmc** omits the redundant moment conditions. The **pdynmc** arguments `inst.stata` and `w.mat.stata` are included to allow for conformity to Stata and to reproduce estimation results.

Empirical example

The functionality of **pdynmc** is illustrated by replicating Arellano and Bond (1991) in a wide sense as we incorporate linear ABov and nonlinear AS moment conditions into the analysis; we also draw comparisons between **pdynmc**, the pgmm (Croissant et al., 2020) function in R-package **plm**, and Stata implementations xtabond2 and xtdpdgmm (Kripfganz, 2019).

Arellano and Bond (1991) employ an unbalanced panel of $n = 140$ firms located in the UK. The dataset spans $T = 9$ time periods and is available from R package **plm**. Arellano and Bond (1991) investigate employment equations and consider the dynamic specification

$$
\begin{aligned}
n_{i,t} =&\, \alpha_1 n_{i,t-1} + \alpha_2 n_{i,t-2} + \\
& \beta_1 w_{i,t} + \beta_2 w_{i,t-1} + \beta_3 k_{i,t} + \beta_4 k_{i,t-1} + \beta_5 k_{i,t-2} + \beta_6 ys_{i,t} + \beta_7 ys_{i,t-1} + \beta_8 ys_{i,t-2} + \\
& \gamma_3 d_3 + \cdots + \gamma_T d_T + \eta_i + \varepsilon_{i,t}, \qquad i = 1, ..., n;\ t = 3, ..., T,
\end{aligned} \tag{11}
$$

where i denotes the firm and t is the time series dimension. The natural logarithm of employment (n) is explained by its first two lags and the further covariates natural logarithm of wage (w), natural logarithm of capital (k), natural logarithm of output (ys), and their lags of order up to one (for w) or two (for k and ys). Variables $d_3, \ldots, d_T$ are time dummies with corresponding coefficients $\gamma_3, \ldots, \gamma_T$; unobserved individual-specific effects are represented by η, and ε is an idiosyncratic remainder component.

We load the dataset and compute logarithms of the four mentioned variables via:

```
data(EmplUK, package = "plm")
dat            <- EmplUK
dat[,c(4:7)]   <- log(dat[,c(4:7)])
names(dat)[4:7] <- c("n", "w", "k", "ys")
```

As GMM estimation with linear and/or nonlinear moment conditions in **pdynmc** allows for arbitrary unbalancedness, we included the functions `data.info` and `strucUPD.plot` to provide an overview of the panel data structure. Both functions require the column name of cross-section (`i.name`) and time series identifier (`t.name`). Using `data.info(dat, i.name = "firm", t.name = "year")` yields

```
Unbalanced panel data set with 1031 rows and the following time period frequencies:
1976 1977 1978 1979 1980 1981 1982 1983 1984
  80  138  140  140  140  140  140   78   35
```

The command `strucUPD.plot(dat,i.name = "firm",t.name = "year")` gives the visual representation of the panel data structure shown in Figure 1. Figure 1 indicates the time periods (compare
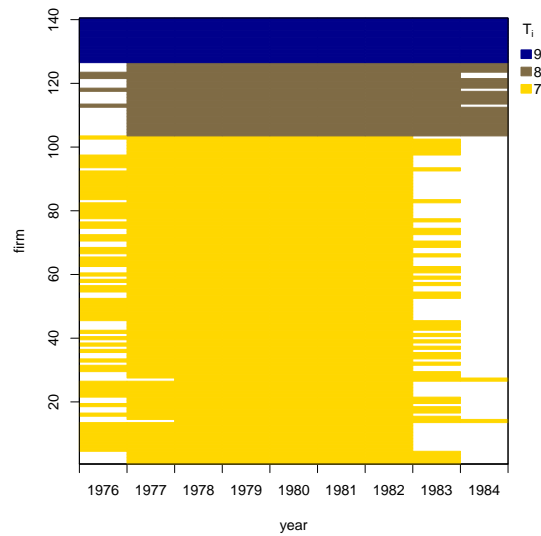

Figure 1: Unbalanced panel structure plot.

abscissa) available for each cross-sectional unit (ordinate). Blank areas represent missing observations. The coloring scheme shows the number of time series units that are available for the corresponding cross-sectional units. We see that the given dataset has already been ordered by the number of time periods available.

The goal of the empirical analysis is to estimate the lag parameters α_1 and α_2 and the coefficients β_j of the $j = 1, \ldots, 8$ further covariates, while controlling for (unobserved) time effects and accounting for unobserved individual-specific heterogeneity. In the following, we first apply **pdynmc** to replicate the original results of Arellano and Bond (1991) that are based on HNR moment conditions only, and introduce the implemented tests. Then, we provide results for adding ABov moment conditions to the analysis. Finally, we discuss results for HNR moment conditions extended by AS moment conditions and apply iterated GMM. All results on estimated coefficients and robust standard errors are summarized in Table 2. Details on employed moment conditions are provided in the table footnotes.

GMM estimation with HNR moment conditions

When reproducing the results in Table 4 on p. 290 of Arellano and Bond (1991) with **pdynmc**, the model structure underlying Equation (11) can be specified and estimated by:

```
m1 <- pdynmc(
  dat = dat, varname.i = "firm", varname.t = "year",
  use.mc.diff = TRUE, use.mc.lev = FALSE, use.mc.nonlin = FALSE,
  include.y = TRUE, varname.y = "n", lagTerms.y = 2,
  fur.con = TRUE, fur.con.diff = TRUE, fur.con.lev = FALSE,
  varname.reg.fur = c("w", "k", "ys"), lagTerms.reg.fur = c(1,2,2),
  include.dum = TRUE, dum.diff = TRUE, dum.lev = FALSE, varname.dum = "year",
  w.mat = "iid.err", std.err = "corrected",
  estimation = "onestep", opt.meth = "none"
)
```

The standard output is accessed via `summary(m1)` and can be found in panel (a) of Table 2. The estimated coefficients reproduce the estimates in Table 4, column (a1) on p. 290 of Arellano and Bond (1991), when one specifies all arguments as stated in this section. Changing the argument `estimation` to `twostep` yields two-step GMM coefficient estimates (the pdynmc-output object is assigned to m2) from Table 4, column (a2) on p. 290 of Arellano and Bond (1991). These results may be found in panel (b) of Table 2. Note that the standard errors presented in column (b) of Table 2 are based on the Windmeijer-correction and deviate from the conventional standard errors reported in Arellano and Bond (1991). Standard errors from the original analysis can be reproduced by setting `std.err = "unadjusted"`.

	(a) 1-Step Estimate HNR only (SE Rob.)	(b) 2-S. Estimate HNR only (SE Rob.)	(c) 2-S. Estimate HNR & ABov (SE Rob.)	(d) 2-S. Estimate HNR & AS (SE Rob.)	(e) Iterated Est. HNR & AS (SE Rob.)
L1.n	0.686***	0.629**	1.103***	1.112***	1.197***
	(0.145)	(0.193)	(0.050)	(0.066)	(0.069)
L2.n	-0.085	-0.065	-0.104*	-0.071	-0.126
	(0.056)	(0.045)	(0.047)	(0.069)	(0.068)
w	-0.608***	-0.526***	-0.448**	-0.417**	-0.219
	(0.178)	(0.155)	(0.149)	(0.153)	(0.127)
L1.w	0.393*	0.311	0.423**	0.413**	0.258
	(0.168)	(0.203)	(0.156)	(0.160)	(0.138)
k	0.357***	0.278***	0.290***	0.309***	0.255***
	(0.059)	(0.073)	(0.050)	(0.053)	(0.056)
L1.k	-0.058	0.014	-0.153*	-0.189**	-0.155*
	(0.073)	(0.092)	(0.067)	(0.068)	(0.077)
L2.k	-0.020	-0.040	-0.137***	-0.154**	-0.156**
	(0.033)	(0.043)	(0.041)	(0.050)	(0.055)
ys	0.609***	0.592***	0.548**	0.582**	0.530**
	(0.173)	(0.173)	(0.194)	(0.178)	(0.183)
L1.ys	-0.711**	-0.566*	-0.666**	-0.624**	-0.379
	(0.232)	(0.261)	(0.221)	(0.216)	(0.223)
L2.ys	0.106	0.101	0.127	0.023	-0.208
	(0.141)	(0.161)	(0.156)	(0.151)	(0.152)
1979	0.010	0.011	0.024*	0.027*	0.031**
	(0.010)	(0.012)	(0.011)	(0.011)	(0.010)
1980	0.022	0.023	0.041*	0.047**	0.053**
	(0.018)	(0.020)	(0.020)	(0.018)	(0.018)
1981	-0.012	-0.021	0.002	0.018	0.026
	(0.030)	(0.033)	(0.034)	(0.030)	(0.030)
1982	-0.027	-0.031	0.018	0.022	0.034
	(0.029)	(0.034)	(0.023)	(0.021)	(0.023)
1983	-0.021	-0.018	0.043*	0.037*	0.041*
	(0.030)	(0.037)	(0.018)	(0.019)	(0.021)
1984	-0.008	-0.023	0.029	0.015	0.021
	(0.031)	(0.037)	(0.022)	(0.022)	(0.024)

(a, b) Equations in differences: $L(2/8).n, D.w, L.D.w, D.k, L.D.k, L2.D.k, D.ys, L.D.ys, L2.D.ys, D.1979 - D.1984$

(c)   Equations in differences: $L(2/8).n, D.w, L.D.w, L2.D.w, D.k, L.D.k, L2.D.k, D.ys, L.D.ys, L2.D.ys, D.1979 - D.1984$

           Equations in levels: $L(1/7).D.n, w, L.w, L2.w, k, L.k, L2.k, ys, L.ys, L2.ys$

(d, e) Equations in differences: $L(2/8).n, u, D.w, L.D.w, L2.D.w, D.k, L.D.k, L2.D.k, D.ys, L.D.ys, L2.D.ys, D.1979 - D.1984$

           Equations in levels: $w, L.w, L2.w, k, L.k, L2.k, ys, L.ys, L2.ys$

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$ (refers to t-test of the null that the coefficient is equal to zero)

Table 2: Estimates in the spirit of Table 4 in Arellano and Bond (1991). Use of $L(a_1/a_2)$ indicates lag transformation by a minimum of a_1 and a maximum of a_2 time periods; D indicates first differences. Table footnotes indicate available instruments and corresponding equations.

The command `mtest.fct(m2,t.order = 2)` is used to perform the test of Arellano and Bond (1991) for second order serial correlation and yields:

```
        Arellano and Bond (1991) serial correlation test of degree 2

data:  2step GMM Estimation; H0: no serial correlation of order 2 in the error terms
normal = -0.37133, p-value = 0.7104
```

The test does not reject the null hypothesis at any plausible significance level and provides no indication that the model specification might be inadequate.

Computing the Hansen *J*-test of overidentifying restrictions by `jtest.fct(m2)` yields:

```
        J-Test of Hansen

data:  2step GMM Estimation; H0: overidentifying restrictions valid
chisq = 31.381, df = 25, p-value = 0.1767
```

As the test does not reject the null hypothesis, there are no indications that the validity of the instruments (i.e., the model assumptions) employed in estimation may be in doubt.

For the Wald test of the null hypothesis that the population parameters of all coefficients included in the model are jointly zero, which is tested by `wald.fct(m2,param = "all")`, we obtain:

```
        Wald test

data:  2step GMM Estimation; H0: all parameters are jointly zero
chisq = 1100, df = 16, p-value < 2.2e-16
```

The test rejects the null hypothesis. Hence, all tests shown here provide no indications that the model in column (b) of Table 2 is misspecified.

Comparing the results to xtabond2 shows that degrees of freedom and *p*-values differ for the latter two tests. We consider 25 degrees of freedom to be the appropriate number in the J-test, as 41 instruments are employed in estimation to obtain 16 coefficient estimates. The latter number (16) is the appropriate number of degrees of freedom in the Wald test. It seems that the function xtabond2 does not correct the degrees of freedom for the number of dummies dropped in estimation[1]. The difference in the *p*-value is due to the differences in the degrees of freedom. Our results are equivalent to the results of pgmm for the overidentifying restrictions test (referred to as "Sargan test" in pgmm).

Using many instruments may have undesirable side effects such as biased coefficient estimates and standard errors; this may result in misleading inference and specification tests (see, e.g., Roodman, 2009). The number of lags of the dependent variable which are used to derive moment conditions can be limited by setting `maxLags.y` (equivalently lags of, for example, endogenous covariates can be limited via `maxLags.reg.end`). Setting `maxLags.y = 4` reduces the number of HNR moment conditions for the GMM estimation above from 27 to 17 and the total number of instruments employed in the estimation from 41 to 31.

GMM estimation with HNR and ABov moment conditions

When the "constant correlated effects" assumption stated in Equation (9) holds, the HNR moment conditions from equations in differences employed in Section GMM estimation with HNR moment conditions can be extended by the ABov moment conditions from equations in levels.

The ABov moment conditions are particularly useful for data generating processes, which are highly persistent (Blundell and Bond, 1998). In this case, identification by the HNR moment conditions from equations in differences may fail and GMM estimation based on HNR moment conditions is documented to possess poor finite sample performance (see, e.g., Blundell and Bond, 1998; Blundell et al., 2001; Bun and Sarafidis, 2015).

In **pdynmc**, the ABov moment conditions from equations in levels can be (additionally) incorporated by:

```
use.mc.lev = TRUE
```

In principle, both time dummies and further covariates can be included in the equations in first differences and the level equations. It is recommended, though, to include the dummies only in one of the equations, as it can be shown that incorporating them in both equations renders one set of

---

[1]Dummies are dropped by the estimation routine in case of high collinearity.

dummies redundant for estimation – while for non-lagged dependent covariates, this equivalence does not hold.[2] We accommodate non-lagged dependent covariates in the levels equations by

```
fur.con.lev = TRUE
```

for all subsequent estimations of this example. The results presented in column (c) of Table 2 are two-step estimates of column (a2) of Table 4 in Arellano and Bond (1991) extended by ABov moment conditions.

Including ABov moment conditions into the analysis leads to substantial changes in the coefficient estimates of the first lag of the dependent variable. Note that the results indicate a markedly higher persistence of employment and render including two lags of the dependent variable questionable (Blundell and Bond, 1998, e.g., estimate a version of the equation which contains only one lag of all covariates). Note that the coefficient estimates of the covariates, besides the first lag of the dependent variable, appear to be similar across estimations.

Equivalent results to column (c) of Table 2 can be obtained from the pgmm function in the **plm**-package. When replicating the results with xtabond2, inst.stata = TRUE in **pdynmc** ensures that results are equivalent.

GMM estimation with HNR and AS moment conditions

Recall that the linear ABov moment conditions from equations in levels encompass the nonlinear AS moment conditions (Blundell and Bond, 1998; a derivation is provided in Fritsch, 2019). Both sets of moment conditions may be useful in GMM estimation when the lag parameter is close to unity and it can be shown that extending the HNR moment conditions by either the ABov or the AS moment conditions may identify the lag parameter – even when individual moment conditions fail to do so (Blundell and Bond, 1998; Bun and Kleibergen, 2021; Gørgens et al., 2019). The ABov moment conditions require the "constant correlated effects" assumption to hold, while the AS moment conditions only require standard assumptions to hold. Therefore, the latter may be useful in situations where the "constant correlated effects" assumption is in doubt and the statistician aims to investigate a highly persistent dynamic process with a structure similar to Equation (3). In **pdynmc**, including nonlinear moment conditions into the analysis is available via:

```
use.mc.nonlin = TRUE
```

When extending the analysis of Arellano and Bond (1991) by the nonlinear AS moment conditions, the results differ substantially from column (b) of Table 2 and are very similar to the coefficient estimates shown in column (c) of Table 2. This indicates high persistence in the employment process that leads to lag parameters not being identified by the HNR moment conditions (Bun and Kleibergen, 2021; Gørgens et al., 2019).

Additionally, we employ iterated GMM via:

```
estimation = "iterative", max.iter = 100, iter.tol = 0.01,
```

Iterated GMM results are shown in column (e) of Table 2. The moment conditions employed are the same as in column (d) of the table. The parameter estimates obtained after 13 steps are relatively similar to those in columns (c)-(d). The ranges of the coefficient estimates across GMM iterations are displayed in Figure 2. This plot is available for two-step and iterated GMM estimates via command plot(m5,type = "coef.range",omit1step = TRUE). Using command plot(m5) yields a scatterplot of fitted values and residuals of a fitted model object, instead.

Figure 2 indicates coefficient estimates at iteration 2 as grey open circles (the first iteration is ignored due to omit1step = TRUE) and estimates at the last iteration as blue diamonds. For the estimates displayed in column (e) of Table 2, we observe that the lag parameters are relatively stable across iterations and resemble the two-step estimates; for the further covariates, larger changes in coefficient estimates across iterations occur for coefficients with larger standard errors (compare w, k, and ys).

As an additional tool to investigate coefficient estimates from iterated GMM, coefficient path plots (compare Hansen and Lee, 2021, Figure 1) are provided. Figure 3 illustrates the path of coefficient estimates for lag parameter α_1 across GMM iterations and is obtained via plot(m5,type = "coef.path",co = "L1.n"). Argument co allows to draw the path(s) of particular coefficient estimates; per default, all coefficients (apart from time dummies) are included in the plot. Approximate 95% confidence bands were added to the plot for the final iteration (available by setting argument add.se.approx = TRUE).

---

[2]Note that this is the case in balanced panels. The results may also not be numerically identical across function calls for different choices of the one-step weighting matrix. For a discussion, see https://www.statalist.org/forums/forum/general-stata-discussion/general/1357268-system-gmm-time-dummies.
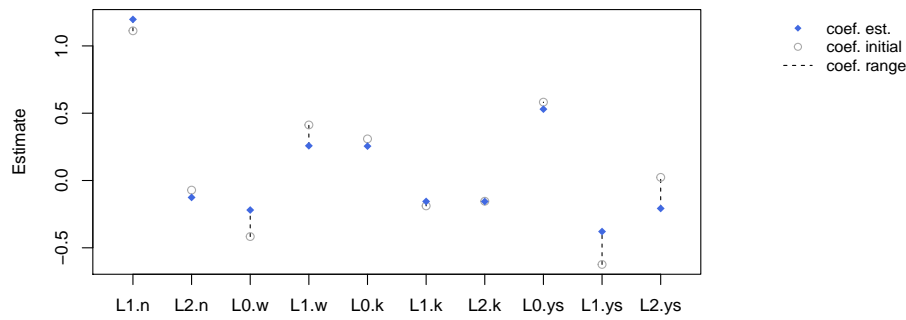
Figure 2: Estimated coefficients and corresponding coefficient ranges during iterated GMM estimation (ordinate) for the covariates (dummy variables excluded). Coefficient estimates at the initial step are displayed as grey open circles and the estimates at the last step as blue diamonds.
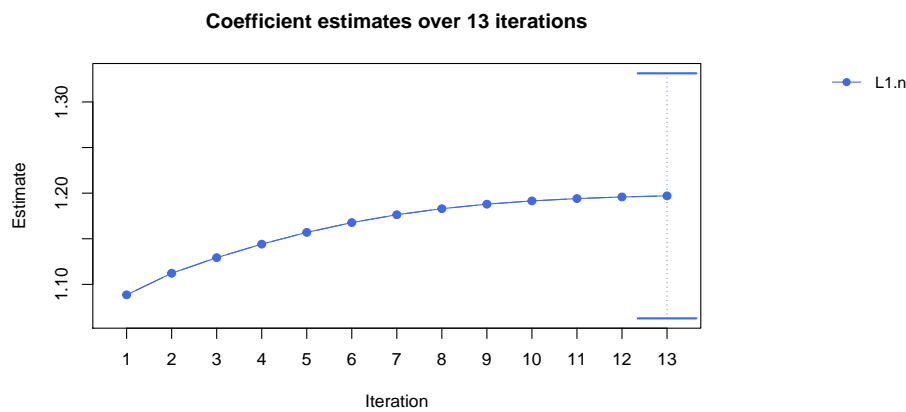

Figure 3: Estimated coefficient path for lag parameter α_1 during iterated GMM estimation and corresponding approximate 95% confidence bands for final iteration.

Overall, the results displayed in columns (c)-(e) of Table 2 suggest that the employment process may be highly persistent and that using only HNR moment conditions may not be sufficient to identify the parameters. In practice, contrasting GMM estimates based on HNR and AS moment conditions with GMM estimates based on HNR and ABov moment conditions can be used as a robustness check of the "constant correlated effects" assumption: When estimates differ, this may cast doubt on the assumption. Here, this is not the case as the results in column (c) are very close to those in (d) and (e).

Conclusion

R-package **pdynmc** provides a function to estimate linear dynamic panel data models based on linear and nonlinear moment conditions. The implementation reflects recent developments in the literature by including iterated GMM and offers a wide variety of configuration options. The package provides the only open source solution for GMM estimation of dynamic panels with linear and nonlinear moment conditions, aligns commercial and noncommercial software, and is implemented to enable the user to exert precise control over all functionality. Additionally, suitable visualizations of panel data structures and ranges and paths of coefficient estimates across GMM iterations are provided.

Functionality of **pdynmc** includes that it allows for general lag structures of the covariates; further controls and external instruments (if available) may also be added. The estimation routine can handle balanced and unbalanced panel data sets and provides one-step-, two-step-, and iterated estimation. Accounting for (unobserved) time-specific effects is possible by including time dummies. Estimation relies on numerical optimization of the GMM objective function. Corresponding closed form solutions are computed – where possible – and stored besides the results from numerical optimization. Different

choices for the weighting matrix, which guides the aggregation of moment conditions in one-step GMM estimation are available. Robust standard errors are available for inference and specification testing. Nonlinear moment conditions provide a robustness check of the frequently employed constant correlated effects assumption.

Acknowledgements

Bibliography

S. C. Ahn and P. Schmidt. Efficient estimation of models for dynamic panel data. *Journal of Econometrics*, 68(1):5–27, 1995. URL https://doi.org/10.1016/0304-4076(94)01641-C. [p1, 2, 3]

T. Anderson and C. Hsiao. Formulation and estimation of dynamic models using panel data. *Journal of Econometrics*, 18(1):47–82, 1982. URL https://doi.org/10.1016/0304-4076(82)90095-1. [p2]

M. Arellano. *Panel Data Econometrics*. Oxford University Press, 2003. URL https://doi.org/10.1093/0199245282.001.0001. [p3, 4]

M. Arellano and S. Bond. Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations. *The Review of Economic Studies*, 58(2):277–297, 1991. URL https://doi.org/10.2307/2297968. [p1, 2, 4, 5, 6, 7, 8, 9, 10]

M. Arellano and O. Bover. Another look at the instrumental variable estimation of error-components models. *Journal of Econometrics*, 68(1):29–51, 1995. URL https://doi.org/10.1016/0304-4076(94)01642-D. [p2, 3]

R. Blundell and S. Bond. Initial conditions and moment restrictions in dynamic panel data models. *Journal of Econometrics*, 87(1):115–143, 1998. URL http://doi.org/10.1016/S0304-4076(98)00009-8. [p9, 10]

R. Blundell, S. Bond, and F. Windmeijer. Estimation in dynamic panel data models: Improving on the performance of the standard GMM estimator. In B. H. Baltagi, T. B. Fomby, and R. C. Hill, editors, *Nonstationary Panels, Panel Cointegration, and Dynamic Panels*, volume 15 of *Advances in Econometrics*, pages 53–91. Emerald Group Publishing, 2001. URL https://doi.org/10.1016/S0731-9053(00)15003-0. [p4, 9]

M. J. G. Bun and F. Kleibergen. Identification robust inference for moments based analysis of linear dynamic panel data models. *Econometric Theory*, forthcoming, 2021. URL https://doi.org/10.2139/ssrn.3847636. [p2, 10]

M. J. G. Bun and V. Sarafidis. Chapter 3 – Dynamic panel data models. In B. H. Baltagi, editor, *The Oxford Handbook of Panel Data*, pages 76–110. Oxford University Press, 2015. URL https://doi.org/10.1093/oxfordhb/9780199940042.013.0003. [p2, 3, 9]

Y. Croissant and G. Millo. *Panel Data Econometrics with R: the plm package*. Wiley, 2019. [p1, 2]

Y. Croissant, G. Millo, and K. Tappe. plm: Linear models for panel data, 2020. URL https://CRAN.R-project.org/package=plm. [p6]

J. A. Doornik, M. Arellano, and S. Bond. *Panel Data estimation using DPD for Ox*, 2012. URL http://www.doornik.com/download.html. [p4]

R. Fletcher. A new approach to variable metric algorithms. *The Computer Journal*, 13(3):317–322, 1970. URL https://doi.org/10.1093/comjnl/13.3.317. [p4]

M. Fritsch. On GMM estimation of linear dynamic panel data models. University of Passau Working Papers in Business Administration B-36-19, University of Passau, 2019. URL https://www.wiwi.uni-passau.de/en/research/working-papers/. [p2, 3, 6, 10]

M. Fritsch, A. A. Y. Pua, and J. Schnurbus. pdynmc – An R-package for estimating linear dynamic panel data models based on nonlinear moment conditions, 2020. URL https://cran.r-project.org/package=pdynmc. [p1, 4]

T. Gørgens, C. Han, and S. Xue. Moment Restrictions and Identification in Linear Dynamic Panel Data Models. *Annals of Economics and Statistics*, (134):149–176, 2019. URL https://doi.org/10.15609/annaeconstat2009.134.0149. [p2, 10]

B. E. Hansen and S. Lee. Inference for Iterated GMM Under Misspecification. *Econometrica*, 89(3): 1419–1447, 2021. URL https://doi.org/10.3982/ECTA16274. [p1, 4, 10]

L. P. Hansen. Large Sample Properties of Generalized Method of Moments Estimators. *Econometrica*, 50(4):1029–1054, 1982. URL https://doi.org/10.2307/1912775. [p4]

D. Holtz-Eakin, K. Newey, Whitney, and H. S. Rosen. Estimating Vector Autoregressions with Panel Data. *Econometrica*, 56(6):1371–1395, 1988. URL https://doi.org/10.2307/1913103. [p2, 3]

C. Hsiao. *Analysis of Panel Data*. Cambridge University Press, 3 edition, 2014. URL https://doi.org/10.1017/CBO9781139839327. [p2]

J. Hwang and Y. Sun. Should we go one step further? An accurate comparison of one-step and two-step procedures in a generalized method of moments framework. *Journal of Econometrics*, 207 (2):381–405, 2018. URL https://doi.org/10.1016/j.jeconom.2018.07.006. [p1]

J. Hwang, B. Kang, and S. Lee. A doubly corrected robust variance estimator for linear GMM. *Journal of Econometrics*, forthcoming, 2021. URL https://doi.org/10.1016/j.jeconom.2020.09.010. [p4]

J. F. Kiviet. Chapter 11 – Judging Contending Estimators by Simulation: Tournaments in Dynamic Panel Data Models. In G. D. A. Phillips and E. Tzavalis, editors, *The Refinement of Econometric Estimation and Test Procedures: Finite Sample and Asymptotic Analysis*, pages 282–318. Cambridge University Press, 2007. URL https://doi.org/10.1017/CBO9780511493157. [p3]

S. Kripfganz. XTDPDGMM: Stata module to perform generalized method of moments estimation of linear dynamic panel data models, 2019. URL http://EconPapers.repec.org/RePEc:boc:bocode:s458395. [p1, 4, 6]

J. C. Nash. *Compact Numerical Methods for Computers: Linear Algebra and Function Minimisation*. CRC Press, 1990. [p4]

J. C. Nash. On Best Practice Optimization Methods in R. *Journal of Statistical Software*, 60(2):1–14, 2014. URL https://doi.org/10.18637/jss.v060.i02. [p4]

J. C. Nash. Provenance of R's Gradient Optimizers. *The R Journal*, 12(1):477–483, 2020. URL https://doi.org/10.32614/RJ-2020-029. [p4]

J. C. Nash and R. Varadhan. Unifying optimization algorithms to aid software system users: optimx for R. *Journal of Statistical Software*, 43(9):1–14, 2011. URL https://doi.org/10.18637/jss.v043.i09. [p4]

M. Pickup, P. Gustafson, D. Cubranic, and G. Evans. OrthoPanels: An R Package for Estimating a Dynamic Panel Model with Fixed Effects Using the Orthogonal Reparameterization Approach. *The R Journal*, 9(1):60–76, 2017. URL https://doi.org/10.32614/RJ-2017-003. [p1]

A. Pua, M. Fritsch, and J. Schnurbus. Large sample properties of an IV estimator based on the Ahn and Schmidt moment conditions. University of Passau Working Papers in Business Administration B-37-19, University of Passau, 2019a. URL https://www.wiwi.uni-passau.de/en/research/working-papers/. [p2]

A. Pua, M. Fritsch, and J. Schnurbus. Practical aspects of using quadratic moment conditions in linear AR(1) panel data models. University of Passau Working Papers in Business Administration B-38-19, University of Passau, 2019b. URL https://www.wiwi.uni-passau.de/en/research/working-papers/. [p2]

D. Roodman. A Note on the Theme of Too Many Instruments. *Oxford Bulletin of Economics and Statistics*, 71(1):135–158, 2009. URL https://doi.org/10.1111/j.1468-0084.2008.00542.x. [p9]

D. Roodman. XTABOND2: Stata module to extend xtabond dynamic panel data estimator, 2018. URL https://econpapers.repec.org/software/bocbocode/s435901.htm. [p6]

M. Sigmund and R. Ferstl. Panel vector autoregression in R with the package panelvar. *The Quarterly Review of Economics and Finance*, page forthcoming, 2019. URL https://doi.org/10.1016/j.qref.2019.01.001. [p1]

StataCorp. *Stata Longitudinal-Data/Panel-Data Reference Manual Release 14*. StataCorp LP, College Station, Texas, 2015. [p1]

F. Windmeijer. A finite sample correction for the variance of linear efficient two-step gmm estimators. *Journal of Econometrics*, 126(1):25–51, 2005. URL https://doi.org/10.1016/j.jeconom.2004.02.005. [p4, 5]

Markus Fritsch
University of Passau
School of Business, Economics, and Information Systems
94032 Passau
Germany
markus.fritsch@uni-passau.de

Andrew Adrian Yu Pua
MOE Key Laboratory of Econometrics
The Wang Yanan Institute for Studies in Economics
Department of Statistics, School of Economics
Fujian Key Laboratory of Statistics
Xiamen University
China
andrewypua@outlook.com

Joachim Schnurbus
University of Passau
School of Business, Economics, and Information Systems
94032 Passau
Germany
joachim.schnurbus@uni-passau.de