The Stata Journal (2016)
16, Number 4, pp. 1013–1038

Quasi–maximum likelihood estimation of linear dynamic short-T panel-data models

Sebastian Kripfganz
University of Exeter Business School
Exeter, UK
S.Kripfganz@exeter.ac.uk

Abstract. In this article, I describe the `xtdpdqml` command for the quasi–maximum likelihood estimation of linear dynamic panel-data models when the time horizon is short and the number of cross-sectional units is large. Based on the theoretical groundwork by Bhargava and Sargan (1983, *Econometrica* 51: 1635–1659) and Hsiao, Pesaran, and Tahmiscioglu (2002, *Journal of Econometrics* 109: 107–150), the marginal distribution of the initial observations is modeled as a function of the observed variables to circumvent a short-T dynamic panel-data bias. Both random-effects and fixed-effects versions are available.

Keywords: st0463, xtdpdqml, dynamic panel data, random effects, fixed effects, short-T bias, quasi–maximum likelihood estimation, initial observations, unbalanced panel data

1 Introduction

The estimation of linear dynamic panel-data models has become increasingly popular in the last decades. When the time horizon is short, ordinary least-squares or generalized least-squares (GLS) estimators for random-effects or fixed-effects models that condition on the initial observations yield biased estimates because of the correlation of the lagged dependent variable with the combined error term.[1] An analytical expression of this bias in fixed-effects models has been obtained by Nickell (1981).

Quasi–maximum likelihood (QML) estimation can circumvent this bias by modeling the unconditional likelihood function instead of conditioning on the initial observations. While this requires additional assumptions about the marginal distribution of the initial observations, the QML estimators are an attractive alternative to other estimation approaches in terms of efficiency and finite-sample performance if all the assumptions are satisfied. Some of those assumptions can be easily tested within the QML framework by means of a likelihood-ratio test if they lead to nested models.

In this article, I describe the new command `xtdpdqml`, which provides an easy-to-use implementation of the QML estimators by Bhargava and Sargan (1983) for the dynamic random-effects model and by Hsiao, Pesaran, and Tahmiscioglu (2002) for the dynamic fixed-effects model. Their estimators are extended to accommodate unbalanced panel

1. In Stata, these least-squares estimators for the random-effects and fixed-effects models are implemented in the command `xtreg`.

data, provided that the sample selection is nonsystematic. Standard errors that are robust to cross-sectional heteroskedasticity are available following Hayakawa and Pesaran (2015). Both QML estimators can be characterized as limited-information maximum-likelihood estimators that are special cases of a structural equation modeling or full-information maximum-likelihood approach with many cross-equation restrictions. For the latter, Williams, Allison, and Moral-Benito (2015) recently presented the similarly named command `xtdpdml`, which builds on Stata's `sem` feature.[2]

When the first two moments of the model are correctly specified, the QML estimators are consistent with potentially sizable efficiency benefits. Yet there is a trade-off between efficiency and robustness. In particular, the QML approach discussed here would turn inconsistent if the explanatory variables (other than the lagged dependent variable) are no longer strictly exogenous with respect to the idiosyncratic error component, or if there is remaining serial correlation that is not captured by the first-order autoregressive term.[3]

In empirical research, using instrumental variables in the context of generalized method of moments (GMM) is the predominant estimation technique to cope with this problem, in part because of the availability of user-friendly estimation commands in standard statistics software. In Stata, the Arellano and Bond (1991) "difference GMM" estimator is implemented in the command `xtabond`, and the "system GMM" extensions by Arellano and Bover (1995) and Blundell and Bond (1998) are implemented in the command `xtdpdsys`. Both commands are wrappers for the more flexible `xtdpd` command, which performs the actual computations. A much respected user-written command with full flexibility and many additional options is `xtabond2`, described in detail by Roodman (2009).

While GMM estimation is very attractive because of its flexibility and ease of implementation, other promising methods remain underrepresented in empirical work. Aside from the QML approach, the bias-correction procedures proposed by Kiviet (1995), Bun and Kiviet (2003), and Everaert and Pozzi (2007), among others, can be a more efficient alternative in dealing with the endogeneity of the lagged dependent variable. Bruno (2005) and De Vos, Everaert, and Ruyssen (2015) provide the user-written implementations `xtlsdvc` and `xtbcfe`, respectively. Both obtain biased estimates first and subsequently remove the bias based on analytical bias expressions or with a bootstrap procedure. In contrast, the QML and GMM approaches are designed to avoid the bias in the first place.

2. Note the missing `q` in the command name `xtdpdml` compared with the `xtdpdqml` command discussed in this article. The names are constructed by combining Stata's `xt` prefix for panel-data commands, `dpd` as an abbreviation for dynamic panel data, and `ml` or `qml` for the full-information maximum-likelihood or the QML method, respectively.
3. While in principal the QML estimators can be extended to include higher-order lags of the dependent variable, this requires additional modeling effort and is not implemented in `xtdpdqml`.

In applied work, the robustness of the estimates obtained with different methods allows an assessment of the reliability of the model's specification assumptions. The new estimation command presented in this article extends the ready-to-use methodological toolkit for the estimation of short-T dynamic panel-data models. It also supports specification testing, in particular the familiar Hausman (1978) test to differentiate between the random-effects and the fixed-effects models.

The remainder of this article is organized as follows. Section 2 outlines the dynamic random-effects and the dynamic fixed-effects models. Section 3 describes the syntax and options of the `xtdpdqml` command, and section 4 does the same for available postestimation commands. An example is discussed in section 5. Section 6 concludes the article. Methodological details are relegated to an online appendix.[4]

2 Dynamic panel-data model

Consider the following linear panel-data model with first-order autoregressive dynamics:

$$y_{it} = \lambda y_{i,t-1} + \mathbf{x}_{it}'\boldsymbol{\beta} + \mathbf{f}_i'\boldsymbol{\gamma} + \epsilon_{it} \qquad \epsilon_{it} = u_i + e_{it} \tag{1}$$

where $\mathbf{x}_{it}$ is a $K_x \times 1$ vector of time-varying variables and $\mathbf{f}_i$ is a $K_f \times 1$ vector of time-invariant variables. The sample is observed for $i = 1, 2, \ldots, N$ cross-sectional units and a short number of $t = 1, 2, \ldots, T_i$ consecutive time periods, with $T_i \geq 2$ possibly different across units but without gaps.[5] The initial observations y_{i0} and $\mathbf{x}_{i0}$ are observed as well. The combined error term ϵ_{it} consists of a time-invariant unit-specific component u_i and an idiosyncratic component e_{it}. The latter is assumed to be independent and identically distributed (i.i.d.) with mean 0 and variance σ_e^2.

2.1 Dynamic random-effects model

Under the random-effects assumption, the unit-specific intercepts u_i are i.i.d. random variables with mean 0 and variance σ_u^2.[6] In particular, they are assumed to be uncorrelated with the exogenous regressors $\mathbf{x}_{it}$ and $\mathbf{f}_i$. Nevertheless, the estimation of model (1) with least-squares techniques conditional on the initial observations is inconsistent when the time horizon is fixed. By construction of the model, the lagged dependent variable $y_{i,t-1}$ is correlated with the time-invariant unit-specific error component u_i, and this is therefore also true for the initial observations y_{i0}. To account for this correlation with a likelihood approach, we need to specify the joint distribution of all observations $\mathbf{y}_i = (y_{i0}, y_{i1}, \ldots, y_{iT_i})'$, conditional on the strictly exogenous regressors $\mathbf{x}_{it}$ and $\mathbf{f}_i$. However, (1) is not sufficient to define the marginal distribution of y_{i0} because of the unobserved $y_{i,-1}$.

4. The online appendix is available at http://www.kripfganz.de.
5. The command `xtdpdqml` automatically drops units with gaps from the estimation sample.
6. The mean 0 assumption is without loss of generality when we include a constant term in the set of time-invariant variables $\mathbf{f}_i$.

Unrestricted initial observations

Instead of assuming y_{i0} to be exogenous, Bhargava and Sargan (1983) advocate the following representation for the initial observations:

$$y_{i0} = \sum_{s=0}^{T^*} \mathbf{x}_{is}' \boldsymbol{\pi}_{x,s} + \mathbf{f}_i' \boldsymbol{\pi}_f + \nu_{i0} \tag{2}$$

where $\boldsymbol{\pi}_{x,s}$ $(s = 1, 2, \ldots, T^*)$ and $\boldsymbol{\pi}_f$ are additional parameter vectors to be estimated. When the panel dataset is balanced—that is, when $T_i = T^*$ for all i—all available observations of the right-hand-side variables can be used in the projection (2). With unbalanced panel datasets, as a computationally straightforward way, I suggest to use as many forward-looking periods as are available for the shortest panel, $T^* = \min(T_i)$, such that (2) is well defined for all units i. With $\text{Var}(\nu_{i0}) = \sigma_0^2$, a suitable parameterization for the covariance between the error terms of the initial observations and the subsequent periods is $\text{Cov}(\nu_{i0}, \epsilon_{it}) = \phi\sigma_0^2$, where both ϕ and σ_0^2 can be treated as free parameters.

Restricted initial observations

Assuming $|\lambda| < 1$ and that the initial observations are generated by the same data-generating process as the remaining observations, we can motivate (2) also by iterating the process continuously backward in time:

$$y_{i0} = \lambda^m y_{i,-m} + \sum_{s=0}^{m-1} \lambda^s \mathbf{x}_{i,-s}' \boldsymbol{\beta} + \frac{1-\lambda^m}{1-\lambda} \mathbf{f}_i' \boldsymbol{\gamma} + \frac{1-\lambda^m}{1-\lambda} u_i + \sum_{s=0}^{m-1} \lambda^s e_{i,-s} \tag{3}$$

If the process for y_{it} started far away in the past, $m \to \infty$, then the first term $\lambda^m y_{i,-m}$ eventually vanishes. Further assuming stationarity of the exogenous regressors, we can project their past and unobserved occurrences, $\mathbf{x}_{i,-s}$ for all $s > 0$, on the observed values of $\mathbf{x}_{is}$ $(s = 0, 1, \ldots, T^*)$ and $\mathbf{f}_i$ to obtain the initial-observations representation proposed by Bhargava and Sargan (1983). Equation (2) can thus be seen as a way to obtain an optimal prediction for the systematic part of y_{i0} conditional upon the observed values of the exogenous variables.[7] Notice that under these additional assumptions the second-to-last term in (3) implies a restriction on the covariance between the initial observations and the unit-specific effects, namely, $\phi\sigma_0^2 = \sigma_u^2/(1-\lambda)$, that can be incorporated into the log-likelihood function to obtain more efficient estimates.

Now reconsider model (1) without exogenous time-varying regressors $\mathbf{x}_{it}$ (while still allowing for time-invariant regressors $\mathbf{f}_i$). Assuming again that the process started in the infinite past and that $|\lambda| < 1$, we obtain

$$y_{i0} = \frac{1}{1-\lambda} \mathbf{f}_i' \boldsymbol{\gamma} + \frac{1}{1-\lambda} u_i + \sum_{s=0}^{\infty} \lambda^s e_{i,-s}$$

7. As $m \to \infty$, the coefficients $\boldsymbol{\pi}_f$ in (2) are equal to $\boldsymbol{\gamma}/(1-\lambda)$ plus a second component that depends on the unknown projection parameters. Unless $\mathbf{f}_i$ does not help to explain the unobserved $\mathbf{x}_{i,-s}$ such that this second component disappears, we can ignore the restriction on the first component and treat $\boldsymbol{\pi}_f$ as a free parameter vector. Similar arguments apply to $\boldsymbol{\pi}_{x,s}$.

A comparison with (2) reveals the parameter restrictions $\boldsymbol{\pi}_f = \boldsymbol{\gamma}/(1-\lambda)$, $\sigma_0^2 = \sigma_u^2/(1-\lambda)^2 + \sigma_e^2/(1-\lambda^2)$, and $\phi\sigma_0^2 = \sigma_u^2/(1-\lambda)$.[8] Because the model with restricted initial observations is nested in the unrestricted model, the validity of these stationarity assumptions can be tested with a likelihood-ratio test.[9]

QML estimation

For balanced panel data, Bhargava and Sargan (1983) set up the log-likelihood function for the system of equations consisting of (1) for all time periods $t \geq 1$ and the initial-observations (2) for period $t = 0$. Under the usual regularity assumptions, the log-likelihood function is well behaved and can be maximized with a gradient-based optimization technique.[10] For such an iterative optimization procedure, appropriate starting values are needed. They can be taken from any initial consistent estimator such as a GMM or minimum distance estimator.[11]

2.2 Dynamic fixed-effects model

The random-effects assumption that rules out any correlation between the unobserved unit-specific effects and the exogenous regressors is often too restrictive. A first step in dealing with this issue is to remove the unit-specific error component by a first-difference transformation,

$$\Delta y_{it} = \lambda \Delta y_{i,t-1} + \Delta \mathbf{x}_{it}' \boldsymbol{\beta} + \Delta e_{it} \tag{4}$$

where the transformed error term now exhibits negative first-order serial dependence, $\mathrm{Cov}(\Delta e_{it}, \Delta e_{i,t-1}) = -\sigma_e^2$. The transformation also removes all time-invariant regressors $\mathbf{f}_i$ from the model.

8. The coefficients $\boldsymbol{\pi}_f$ and the variance σ_0^2 are no longer confounded by projections of unobserved on observed variables. Compare with Hsiao, Pesaran, and Tahmiscioglu (2002) for the autoregressive model with a constant term only. Conceptually, time-invariant regressors are no different than the constant term here. With the `xtdpdqml` command, the restrictions can be imposed by the option `stationary`; see section 3.2. However, the assumption $|\lambda| < 1$ is not enforced, which may lead to contradictory results.
9. The stationarity assumptions are sufficient for (2) to be a valid initial condition for consistent estimation of the parameters of interest. However, (2) without the parameter restrictions also remains valid under alternative assumptions, for example, if $\lambda \geq 1$ but the process was initialized in the finite past; see Bhargava and Sargan (1983).
10. Maximization algorithms supported by `xtdpdqml` are Stata's modified Newton–Raphson algorithm, `technique(nr)`; the Davidon–Fletcher–Powell algorithm, `technique(dfp)`; the Broyden–Fletcher–Goldfarb–Shanno algorithm, `technique(bfgs)`; and combinations of them. See Gould, Pitblado, and Poi (2010) for details. Further options for controlling the optimization procedure are available; see section 3.2. The unrestricted and restricted log-likelihood functions and their analytical first-order and second-order derivatives are documented for unbalanced panel data in the online appendix.
11. By default, the `xtdpdqml` command uses GMM estimates for the coefficients λ, $\boldsymbol{\beta}$, and $\boldsymbol{\gamma}$. Starting values for the initial-observations parameters are obtained from a separate ordinary least-squares estimation, and starting values for the variance parameters σ_u^2, σ_e^2, σ_0^2, and ϕ are calculated based on the estimated residuals. Alternative starting values may be specified with the `from()` and `initval()` options. See section 3.2 and the online appendix for further details.

Maximum likelihood estimation of the transformed model (4) conditional on the initial observations under i.i.d. normally distributed disturbances e_{it} is equivalent to GLS estimation. As demonstrated by Bun and Kiviet (2006), the GLS estimator is invariant to the model transformation in balanced panels and equals the least-squares dummy-variables estimator. Because of the correlation of the initial observations Δy_{i1} with the transformed error term, such a conditional likelihood approach is inconsistent when the time horizon is fixed. Again, specifying the joint distribution of $\Delta \mathbf{y}_i = (\Delta y_{i1}, \Delta y_{i2}, \ldots, \Delta y_{iT_i})'$ conditional on the strictly exogenous regressors $\mathbf{x}_{it}$ can address this problem.[12]

Unrestricted initial observations

Similarly to the random-effects model, Hsiao, Pesaran, and Tahmiscioglu (2002) propose the following feasible representation for the initial observations of the transformed model:

$$\Delta y_{i1} = b + \sum_{s=1}^{T^*} \Delta \mathbf{x}_{is}' \boldsymbol{\pi}_s + \nu_{i1} \tag{5}$$

with $T^* = \min(T_i)$ as before. A useful parameterization for the variance of the initial-observations projection error turns out to be $\mathrm{Var}(\nu_{i1}) = \omega \sigma_e^2$, where ω can be treated as a free parameter. The projection error further satisfies the properties $\mathrm{Cov}(\nu_{i1}, \Delta e_{i2}) = -\sigma_e^2$ and $\mathrm{Cov}(\nu_{i1}, \Delta e_{it}) = 0$ for $t = 3, 4, \ldots, T_i$.

Restricted initial observations

To motivate (5), we can apply the same idea as in the random-effects model. Assuming that the initial observations are generated by the same data-generating process as the subsequent observations, we can iterate the process continuously backward to obtain

$$\Delta y_{i1} = \lambda^m \Delta y_{i,1-m} + \sum_{s=0}^{m-1} \lambda^s \Delta \mathbf{x}_{i,1-s}' \boldsymbol{\beta} + \sum_{s=0}^{m-1} \lambda^s \Delta e_{i,1-s} \tag{6}$$

Further assuming that the strictly exogenous regressors $\mathbf{x}_{it}$ are trend or first-difference stationary, Hsiao, Pesaran, and Tahmiscioglu (2002) project the unobserved terms, $\Delta \mathbf{x}_{i,1-s}$ for all $s > 0$, on the current and observed realizations of the transformed regressors, $\Delta \mathbf{x}_{is}$, $s = 1, 2, \ldots, T^*$. Under the stationarity assumption, $|\lambda| < 1$ and $m \to \infty$, the first term $\lambda^m \Delta y_{i,1-m}$ vanishes. The resulting initial-observations representation is again (5) but with the restriction $b = 0$, provided that in addition the exogenous regressors are stationary in levels (or integrated of order 1 without drift).

12. In contrast to the argument by Hsiao, Pesaran, and Tahmiscioglu (2002), the estimator is inconsistent if the regressors are weakly exogenous (predetermined) because of the serial correlation of the transformed errors.

Finally, consider a situation without time-varying regressors $\mathbf{x}_{it}$. With $|\lambda| < 1$ and $m \to \infty$, (6) simplifies to

$$\Delta y_{i1} = \sum_{s=0}^{\infty} \lambda^s \Delta e_{i,1-s}$$

A comparison with (5) reveals the restrictions $b = 0$ and $\omega = 2/(1+\lambda)$.[13]

QML estimation

Hsiao, Pesaran, and Tahmiscioglu (2002) provide the log-likelihood function in the case of balanced panel data for the system of equations formed by (4) for the time periods $t \geq 2$ and (2) for $t = 1$. It can again be maximized with an iterative procedure. As in the random-effects model, appropriate starting values may be obtained from initial consistent estimates.[14]

3 The xtdpdqml command

The `xtdpdqml` command has standard Stata syntax known from other estimation commands. The lagged dependent variable is added automatically to the set of regressors.[15] Several options are available to specify the precise model and to control the optimization process. The default is to fit a fixed-effects model.

13. With `xtdpdqml`, the restrictions on b and ω can be imposed by the option `stationary`; see section 3.2. However, the assumption $|\lambda| < 1$ is not enforced. Hsiao, Pesaran, and Tahmiscioglu (2002) also consider the alternative assumption that the process started from a finite period in the past with identical expected changes in the initial endowments across all units i and without requiring that $|\lambda| < 1$. In this case, the intercept b is still allowed to be nonzero and $\omega = 2(1+\lambda^{2m-1})/(1+\lambda)$ can be treated as a free parameter as long as m is unknown and identical for all i.
14. The starting values used by the `xtdpdqml` command are obtained in a similar way to the random-effects model. Details as well as analytical expressions of the unrestricted and restricted log-likelihood functions and their respective derivatives for the case of unbalanced panel data are documented in the online appendix.
15. `xtdpdqml` does not support higher-order autoregressive dynamics. Including distributed lags of the exogenous regressors is straightforward by using Stata's time-series lag operator `L.`.

3.1 Syntax

Fixed-effects model

`xtdpdqml` *depvar* [*indepvars*] [*if*] [*in*] [, `fe` *projectionopt_fe* `stationary` `noconstant` `vce(`*vcetype*`)` `mlparams` `level(`#`)` `coeflegend` `noheader` `notable` `first` `neq(`#`)` *display_options* `from(`*init_specs*`)` `storeinit(`*name*`)` `initval(`*numlist*`)` `inititer(`#`)` `concentration` `method(`*method*`)` *maximize_options*]

where *projectionopt_fe* is

`projection(`*varlist* [, `leads(`#`)` `nodifference` `omit`]`)`

Random-effects model

`xtdpdqml` *depvar* [*indepvars*] [*if*] [*in*], `re` [*projectionopt_re* `stationary` `noeffects` `noconstant` `vce(`*vcetype*`)` `mlparams` `level(`#`)` `coeflegend` `noheader` `notable` `first` `neq(`#`)` *display_options* `from(`*init_specs*`)` `storeinit(`*name*`)` `initval(`*numlist*`)` `method(`*method*`)` *maximize_options*]

where *projectionopt_re* is

`projection(`*varlist* [, `leads(`#`)` `omit`]`)`

3.2 Options[16]

Model

`projection(`*varlist*[, `leads(`#`)` `nodifference` `omit`]`)` specifies the exogenous variables that are used in the initial-observations projection. `leads(`#`)` restricts the number of leads. The default is `leads(.)`, which means that all available leads are used. In the fixed-effects model, first differences of *varlist* are used unless `nodifference` is specified. By default, all *indepvars* are used unless *varlist* is excluded with `omit`. You may specify as many sets of projection variables as you need.[17]

`stationary` assumes that the process of *depvar* started in the infinite past, that the autoregressive coefficient is less than unity in absolute value (which is not enforced),

16. Further information related to the available options can be found in the online appendix.
17. The `projection()` option specifies the right-hand-side variables of (2) or (5), respectively, as discussed in section 2. The default specifications are those suggested by Bhargava and Sargan (1983) or Hsiao, Pesaran, and Tahmiscioglu (2002).

and that all *indepvars* are stationary as well (in first differences if a fixed-effects model is fit). As a consequence, the initial-observations parameters are restricted to equal their long-run values if there are no time-varying *indepvars*, and the constant term in the initial-observations equation is restricted to 0 (unless a random-effects model with constant term is fit). By default, none of the parameter restrictions are imposed.[18]

`noeffects` restricts the variance of the unit-specific error component in the random-effects model to be 0.

`noconstant`; see [R] **estimation options**.[19]

SE/Robust

`vce(`*vcetype*`)` specifies the type of standard error reported, which includes types that are robust to some kinds of misspecification (`robust`) and that are derived from asymptotic theory (`oim`, `opg`); see [R] ***vce_option***.

`vce(oim)`, the default, uses the observed information matrix.

`vce(opg)` uses the sum of the outer product of the gradient vectors. This option is seldom used.

`vce(robust)` uses the sandwich estimator.

Reporting

`mlparams` reports all QML parameter estimates including the model coefficients, the initial-observations coefficients, and the variance parameters. By default, only the model coefficients are reported.

`level(`#`)`; see [R] **estimation options**.

`coeflegend`; see [R] **estimation options**.

`noheader` suppresses display of the header above the coefficient table that displays the number of observations.

`notable` suppresses display of the coefficient table.

`first` in combination with `mlparams` displays a coefficient table reporting results for the first equation only and makes it appear as if only one equation was estimated.

`neq(`#`)` in combination with `mlparams` displays a coefficient table reporting results for the first # equations. `neq(1)` is equivalent to `first`. `neq(2)` displays the model coefficients and the initial-observations coefficients.

18. In the random-effects model, the `stationary` option enforces restrictions on the parameters $\boldsymbol{\pi}_f$, σ_0^2, and ϕ. In the fixed-effects model, it enforces restrictions on the parameters b and ω. The respective restrictions depend on the inclusion of time-varying exogenous variables in the model; see section 2.

19. If the constant is suppressed by the `noconstant` option, an intercept is still included in the marginal distribution of the initial observations unless the `stationary` option is specified as well.

display_options: `noci`, `nopvalues`, `noomitted`, `vsquish`, `noemptycells`, `baselevels`, `allbaselevels`, `nofvlabel`, `fvwrap(#)`, `fvwrapon(`*style*`)`, `cformat(%`*fmt*`)`, `pformat(%`*fmt*`)`, `sformat(%`*fmt*`)`, and `nolstretch`; see [R] **estimation options**.

Maximization

`from(`*init_specs*`)` specifies initial values for the coefficients; see [R] **maximize**. By default, initial values are taken from GMM estimation; see [XT] **xtdpd**.[20]

`storeinit(`*name*`)` stores the initial GMM estimation results; see [R] **estimates store**.

`initval(`*numlist*`)` specifies initial values for the variance parameters. In the fixed-effects model, at most two numbers are allowed. The first entry refers to the variance of the idiosyncratic error component, σ_e^2, and the second entry refers to the initial-observations variance relative to that of the idiosyncratic component, ω. By default, the first parameter is computed from the residuals given the initial coefficient values, and the last parameter is computed from the first-order condition of the maximization problem given all other parameters. In the random-effects model, at most four numbers are allowed. The first entry refers to the variance of the unit-specific error component, σ_u^2; the second entry refers to the variance of the idiosyncratic error component, σ_e^2; the third entry refers to the initial-observations variance, σ_0^2; and the fourth entry refers to the covariance of the initial observations with the unit-specific error component relative to the initial-observations variance, ϕ. Missing values are allowed to request the default initialization.[21] This option is seldom used.

`inititer(#)` specifies the number of iterations used to update the initial values before maximizing the log-likelihood function of the transformed fixed-effects model. `inititer(0)`, the default, uses the initial values for the coefficients and variance parameters as specified with the `from()` and `initval()` options. `inititer(1)` starts the maximization with the minimum distance estimates given the estimate of the initial-observations variance parameter from the previous step. From the second iteration onward, the analytical first-order condition for the initial-observations variance parameter is evaluated at the parameter values from the previous iteration step. Subsequently, new minimum distance estimates are obtained for the other parameters given the updated value of the initial-observations variance parameter.

`concentration` specifies that the concentrated log-likelihood function of the transformed fixed-effects model with the initial-observations variance as single parameter should be maximized. All other parameter estimates are obtained from the analytical first-order conditions given the optimal value of the initial-observations variance parameter. By default, maximization is done over all parameters simultaneously.

20. Initial one-step GMM estimates are obtained with GMM-type instruments for the lagged dependent variable, as proposed by Arellano and Bond (1991), and with standard instruments for the strictly exogenous regressors in the first-differenced equation. If applicable, standard instruments for time-invariant regressors are added to the level equation in the random-effects model, as suggested by Arellano and Bover (1995).
21. See the online appendix for the formula used to compute the default initial values.

A concentrated log-likelihood function is not available with the `stationary` option when the model is a pure autoregressive process without additional *indepvars*.

`method(`*method*`)` specifies the evaluator method for the log-likelihood function, where *method* is one of `d0`, `d1`, or `d2` (or the respective long form); see [R] **ml**. The default is `method(d2)`. This option is seldom used.

maximize_options: `technique(`*algorithm_spec*`)`, `iterate(#)`, `[no]log`, `showstep`, `showtolerance`, `tolerance(#)`, `ltolerance(#)`, `nrtolerance(#)`, and `nonrtolerance`; see [R] **maximize**. These options are seldom used. Supported *algorithm_specs* are `nr`, `dfp`, and `bfgs` (and combinations). `iterate(0)` may be used to evaluate the log-likelihood function at the initial parameter values.

3.3 Stored results

`xtdpdqml` stores the following in `e()`:

Scalars			
`e(N)`	number of observations	`e(k_eq)`	number of equations in `e(b)`
`e(N_g)`	number of groups	`e(ll)`	log likelihood
`e(g_min)`	smallest group size	`e(rank)`	rank of the Hessian matrix from the `ml` optimization
`e(g_avg)`	average group size		
`e(g_max)`	largest group size	`e(ic)`	number of iterations
`e(k_aux)`	number of ancillary parameters in `e(b)`	`e(converged)`	`1` if converged, `0` otherwise
		`e(stationary)`	`1` if option `stationary` specified
Macros			
`e(cmd)`	`xtdpdqml`	`e(ivar)`	variable denoting groups
`e(cmdline)`	command as typed	`e(tvar)`	variable denoting time
`e(depvar)`	name of dependent variable	`e(properties)`	`b V`
`e(vce)`	*vcetype* specified in `vce()`	`e(predict)`	program used to implement `predict`
`e(vcetype)`	title used to label Std. Err.		
`e(ml_method)`	type of `ml` method	`e(model)`	`re` or `fe`
Matrices			
`e(b)`	coefficient vector	`e(V)`	variance–covariance matrix of the estimators
`e(ilog)`	iteration log (up to 20 iterations)	`e(V_modelbased)`	model-based variance; not always saved
`e(gradient)`	gradient vector		
Functions			
`e(sample)`	marks estimation sample		

4 Postestimation commands

`xtdpdqml` supports many postestimation commands, including `hausman`, `lrtest`, `nlcom`, `predict`, and `test`.[22] Predictions are obtained in a similar way to Stata's `xtreg` command. In addition, `predict` supports the computation of equation-level scores. As a consequence, the `suest` command works after `xtdpdqml`, for example, to combine estimation results from the random-effects and the fixed-effects models to perform a generalized Hausman test.[23]

22. See `help xtdpdqml postestimation` for an extended list.
23. `suest` requires `xtdpdqml` to be used with option `mlparams`; see section 5 for an example.

4.1 Syntax for predict

`predict` [*type*] *newvar* [*if*] [*in*] [, `xb`|`stdp`|`ue`|`xbu`|`u`|`e` `equation(`*eqno*`)`]

`predict` [*type*] {*stub**|*newvar1* ... *newvarq*} [*if*] [*in*], `scores`
  [`equation(`*eqno*`)`]

4.2 Options for predict

`xb`, the default, calculates the linear prediction from the fitted model; see [R] **predict**. After `xtdpdqml, fe mlparams`, it calculates the linear prediction from the first-differenced model.

`stdp` calculates the standard error of the linear prediction; see [R] **predict**.

`ue` calculates the prediction of $u_i + e_{it}$, the combined residual; see [XT] **xtreg postestimation**. This option is not available after `xtdpdqml, fe mlparams`.

`xbu` calculates the linear prediction including the unit-specific error component; see [XT] **xtreg postestimation**. This option is not available after `xtdpdqml, fe mlparams`.

`u` calculates the prediction of u_i, the estimated unit-specific error component; see [XT] **xtreg postestimation**. This option is not available after `xtdpdqml, fe mlparams`.

`e` calculates the prediction of e_{it}; see [XT] **xtreg postestimation**. After `xtdpdqml, fe mlparams`, it calculates the prediction of Δe_{it}, the first-differenced residual.

`scores` calculates the equation-level score variables; see [R] **predict**. This is the derivative of the log-likelihood function with respect to the linear prediction. Ancillary parameters make up separate equations. This option is available only after `xtdpdqml, mlparams` without the `stationary` option.

`equation(`*eqno*`)` specifies the equation to which you are referring; see [R] **predict**.

5 Example

Let us now consider an example based on `abdata.dta`, which contains unbalanced labor demand data for 140 companies in the United Kingdom during the period 1976–1984:

```
. webuse abdata
```

It is the dataset used by Arellano and Bond (1991) in their influential article on GMM estimation of dynamic panel-data models. They estimate employment equations to explain the logarithm of the number of employees (`n`). Strictly exogenous explanatory variables are the real wage (`w`), the gross capital stock (`k`), and the industry output (ys). For a concise presentation, I ignore the last variable, which Arellano and Bond find to be statistically insignificant in their analysis, and I also refrain from including distributed lags of the exogenous regressors. After losing one observation because of the lagged dependent variable, we can include time dummies for the years from 1978 to 1984. As a first step, let us estimate a dynamic fixed-effects model with the Hsiao, Pesaran, and Tahmiscioglu (2002) QML estimator, the default of `xtdpdqml`:

```
. xtdpdqml n w k yr1978-yr1984
Quasi-maximum likelihood estimation
Iteration 0:   f(p) =  488.47979
Iteration 1:   f(p) =  560.74524
Iteration 2:   f(p) =  670.55622
Iteration 3:   f(p) =  693.49986
Iteration 4:   f(p) =   694.4841
Iteration 5:   f(p) =  694.49228
Iteration 6:   f(p) =  694.49228
Group variable: id                            Number of obs         =       891
Time variable: year                           Number of groups      =       140
Fixed effects                                 Obs per group:    min =         6
                                                                avg =  6.364286
(Estimation in first differences)                               max =         8

           n       Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]

           n
         L1.    .7181159   .0349792    20.53   0.000     .6495579    .7866738

           w   -.4210157   .0512701    -8.21   0.000    -.5215034   -.3205281
           k    .2487324   .0255407     9.74   0.000     .1986736    .2987911
      yr1978   -.0214489   .0149487    -1.43   0.151    -.0507478      .00785
      yr1979   -.0319754   .0149372    -2.14   0.032    -.0612518   -.0026991
      yr1980   -.0637126   .0148821    -4.28   0.000     -.092881   -.0345441
      yr1981   -.1130657   .0150739    -7.50   0.000      -.14261   -.0835213
      yr1982   -.0844508   .0160798    -5.25   0.000    -.1159666    -.052935
      yr1983   -.0461928   .0197008    -2.34   0.019    -.0848057   -.0075798
      yr1984   -.0115354   .0241271    -0.48   0.633    -.0588236    .0357528
       _cons     1.74826   .1705756    10.25   0.000     1.413938    2.082582
```

The results are reported for the levels (1), even though the actual estimation is performed on the first-differenced (4).[24] For clarity of the main results, the default output table does not include the additional coefficients from the initial-observations projection (5) and the ancillary variance parameters. We can display the whole set of parameter estimates with the `mlparams` option, suppressing for convenience the iteration log with option `nolog`:

24. This is in line with the "difference GMM" estimation command `xtabond`. While the first-difference transformation removes all time-invariant variables, `xtdpdqml` still reports a constant term for the fixed-effects model in levels unless the `noconstant` or `mlparams` option is specified. It is obtained with the two-stage approach proposed by Kripfganz and Schwarz (2015). The first-stage residuals from the untransformed equation, $y_{it} - \widehat{\lambda} y_{i,t-1} - \mathbf{x}_{it}'\widehat{\boldsymbol{\beta}}$, are regressed on a constant term, and the standard errors are appropriately corrected to account for the first-stage estimation error.

```
. xtdpdqml n w k yr1978-yr1984, mlparams nolog
Quasi-maximum likelihood estimation
Group variable: id                            Number of obs        =       891
Time variable: year                           Number of groups     =       140
Fixed effects                                 Obs per group:    min =         6
                                                                avg =  6.364286
                                                                max =         8

------------------------------------------------------------------------------
         D.n |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
_model       |
           n |
         LD. |   .7181159   .0349792    20.53   0.000     .6495579    .7866738
             |
           w |
         D1. |  -.4210157   .0512701    -8.21   0.000    -.5215034   -.3205281
             |
           k |
         D1. |   .2487324   .0255407     9.74   0.000     .1986736    .2987911
             |
      yr1978 |
         D1. |  -.0214489   .0149487    -1.43   0.151    -.0507478      .00785
             |
      yr1979 |
         D1. |  -.0319754   .0149372    -2.14   0.032    -.0612518   -.0026991
             |
      yr1980 |
         D1. |  -.0637126   .0148821    -4.28   0.000     -.092881   -.0345441
             |
      yr1981 |
         D1. |  -.1130657   .0150739    -7.50   0.000      -.14261   -.0835213
             |
      yr1982 |
         D1. |  -.0844508   .0160798    -5.25   0.000    -.1159666    -.052935
             |
      yr1983 |
         D1. |  -.0461928   .0197008    -2.34   0.019    -.0848057   -.0075798
             |
      yr1984 |
         D1. |  -.0115354   .0241271    -0.48   0.633    -.0588236    .0357528
-------------+----------------------------------------------------------------
_initobs     |
           w |
         D1. |   .1745629   .0835193     2.09   0.037      .010868    .3382578
         FD. |   .4866594   .1160984     4.19   0.000     .2591107     .714208
        F2D. |    .234992   .0921914     2.55   0.011     .0543001    .4156838
        F3D. |    .180422   .0831649     2.17   0.030     .0174218    .3434222
        F4D. |   .1587507   .0822884     1.93   0.054    -.0025316    .3200329
        F5D. |   .1828358   .0801948     2.28   0.023      .025657    .3400147
             |
           k |
         D1. |   .2516903   .0514379     4.89   0.000     .1508739    .3525068
         FD. |  -.0759983   .0442764    -1.72   0.086    -.1627784    .0107819
        F2D. |   .0345647   .0402481     0.86   0.390    -.0443201    .1134496
        F3D. |   .0426643   .0416536     1.02   0.306    -.0389754    .1243039
        F4D. |   .0180357   .0354471     0.51   0.611    -.0514394    .0875108
        F5D. |   .1373772   .0420249     3.27   0.001     .0550099    .2197445
             |
```

```
     yr1978 |
        D1. |   .0472505   .0347851     1.36   0.174    -.0209269     .115428
        FD. |   .0336196   .0205327     1.64   0.102    -.0066237     .073863
            |
      _cons |   .0034106   .0211468     0.16   0.872    -.0380363    .0448575
------------+----------------------------------------------------------------
  /_sigma2e |   .0107403   .0005952                      .0095737     .011907
    /_omega |   1.219196   .0690326                      1.083894    1.354497
-----------------------------------------------------------------------------

. estimates store fe
```

The first equation in the output table, labeled `_model`, reports again the main coefficients of interest but now highlights the first-difference transformation. The second equation, labeled `_initobs`, contains the initial-observations coefficients. The last two parameters in the table refer to the estimates of the variance parameters σ_e^2 and ω. In balanced panels, all the time dummies would have been omitted automatically from the `_initobs` equation because of perfect collinearity with the constant term. In unbalanced panels, as in the present case, the initial period might differ across units and some time dummies are retained to account for differences in the initialization.

We observe that the first three coefficients in the `_model` equation are highly statistically significant and the same is true for the majority of the time dummies. The coefficient of the lagged dependent variable is well within the stationarity region, $|\lambda| < 1$. Under some additional assumptions, this could imply that the initial-observations intercept vanishes. We indeed observe that the constant term in the `_initobs` equation is not statistically significant. To obtain more-efficient estimates, we can remove this intercept with the `stationary` option:[25]

```
. xtdpdqml n w k yr1978-yr1984, stationary mlparams nolog
Quasi-maximum likelihood estimation
Group variable: id                          Number of obs        =       891
Time variable: year                         Number of groups     =       140
Fixed effects                               Obs per group:   min =         6
                                                             avg =  6.364286
                                                             max =         8

-----------------------------------------------------------------------------
        D.n |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
------------+----------------------------------------------------------------
_model      |
          n |
        LD. |   .7175702   .0347616    20.64   0.000     .6494386    .7857017
            |
          w |
        D1. |  -.4219682   .0509203    -8.29   0.000    -.5217701   -.3221662
            |
          k |
        D1. |   .2493911   .0251776     9.91   0.000     .2000439    .2987384
            |
```

25. That is, $b = 0$ in (5) under the assumptions of $|\lambda| < 1$, an initialization in the infinite past, and trend or first-difference stationarity of the exogenous regressors `w` and `k`. Because of the presence of the latter, the variance parameter ω remains unrestricted. See the discussion about restricted initial observations in section 2.2.

yr1978						
D1.	-.0212959	.0149167	-1.43	0.153	-.0505321	.0079404
yr1979						
D1.	-.0317929	.0148925	-2.13	0.033	-.0609817	-.0026041
yr1980						
D1.	-.0633101	.0146697	-4.32	0.000	-.0920621	-.0345581
yr1981						
D1.	-.1125881	.0147782	-7.62	0.000	-.141553	-.0836233
yr1982						
D1.	-.0839164	.0157373	-5.33	0.000	-.1147609	-.053072
yr1983						
D1.	-.0455604	.0193118	-2.36	0.018	-.0834109	-.0077099
yr1984						
D1.	-.0107753	.0236674	-0.46	0.649	-.0571625	.0356119
_initobs						
w						
D1.	.1734465	.0833066	2.08	0.037	.0101686	.3367244
FD.	.4915282	.1122137	4.38	0.000	.2715935	.711463
F2D.	.2351962	.0922567	2.55	0.011	.0543763	.416016
F3D.	.1847706	.0787435	2.35	0.019	.0304362	.3391051
F4D.	.1623383	.0793019	2.05	0.041	.0069094	.3177673
F5D.	.1883984	.0724927	2.60	0.009	.0463153	.3304815
k						
D1.	.252992	.0508592	4.97	0.000	.1533099	.3526741
FD.	-.0768106	.0440244	-1.74	0.081	-.1630968	.0094757
F2D.	.0344116	.0402711	0.85	0.393	-.0445184	.1133416
F3D.	.0410705	.0404996	1.01	0.311	-.0383073	.1204483
F4D.	.0168102	.0346589	0.49	0.628	-.05112	.0847404
F5D.	.13622	.0414449	3.29	0.001	.0549895	.2174506
yr1978						
D1.	.0515849	.0221159	2.33	0.020	.0082386	.0949312
FD.	.035909	.0148529	2.42	0.016	.0067979	.0650202
/_sigma2e	.0107368	.0005943			.009572	.0119015
/_omega	1.22007	.0688652			1.085097	1.355044

```
. estimates store fe_s
```

The remaining coefficients changed only slightly, but the two time effects in the `_initobs` equation turned statistically significant at the 5% level. This does not invalidate our assumptions. All it means is that we should control for different starting points in the observed sample. We have stored the results from the previous two estimations under the names `fe` and `fe_s`, respectively, which we can now use to double check the validity of the imposed restriction with a likelihood-ratio test:

```
. lrtest fe_s fe
Likelihood-ratio test                                 LR chi2(1)  =      0.03
(Assumption: fe_s nested in fe)                       Prob > chi2 =    0.8720
```

The test does not reject the restriction on the initial-observations intercept such that we can retain the stationarity assumption in the further analysis. If the time effects in the _initobs equation were jointly insignificant, we could also exclude them by specifying the option `projection(yr*, omit)`:[26]

```
. xtdpdqml n w k yr1978-yr1984, stationary projection(yr*, omit) mlparams nolog
  (output omitted)
. lrtest fe
Likelihood-ratio test                                 LR chi2(3)  =      6.29
(Assumption: . nested in fe)                          Prob > chi2 =    0.0983
```

Based on the conservative significance level of 10%, this likelihood-ratio test suggests not to exclude the time dummies jointly with the intercept from the initial-observations equation. As an asymptotically equivalent test, the Wald test on joint insignificance of the respective coefficients in the unrestricted model yields the same conclusion:

```
. estimates restore fe
(results fe are active now)
. test [_initobs]: D.yr1978 FD.yr1978 _cons
 ( 1)  [_initobs]D.yr1978 = 0
 ( 2)  [_initobs]FD.yr1978 = 0
 ( 3)  [_initobs]_cons = 0
           chi2(  3) =    6.36
         Prob > chi2 =    0.0955
```

The `projection()` option can also be used to restrict the number of time leads of the exogenous regressors in (5). By default, all current and future observations (up to the shortest time length in unbalanced panels) are used as separate variables in the initial-observations projection. With a large time dimension, the number of corresponding coefficients becomes large as well. This problem is aggravated with an increasing number of exogenous time-varying regressors. The following example illustrates how to use only contemporaneous values of the first-differenced regressors:[27]

```
. xtdpdqml n w k yr1978-yr1984, stationary projection(w k, leads(0)) mlparams
  (output omitted)
. lrtest fe_s
Likelihood-ratio test                                 LR chi2(10) =     42.46
(Assumption: . nested in fe_s)                        Prob > chi2 =    0.0000
```

In the present case, the number of parameters is reasonably small, and the likelihood-ratio test clearly rejects this restricted-model version. Alternatively, we might want to use all the available levels of the exogenous variables instead of their first differences

26. This excludes the time dummies only from the `_initobs` equation and not from the `_model` equation.
27. An alternative might be to use the scores of a principal component analysis applied on the projection variables in (5). Bontempi and Mammi (2015) suggest such a strategy to reduce the instrument count for GMM estimators, and it is also implemented in the `xtabond2` command by Roodman (2009). To be used with `xtdpdqml`, these scores would have to be computed separately beforehand—for example, with the `pca2` command by Bontempi and Mammi (2015)—and could then be supplied with the `projection()` option. The credit for this idea goes to an anonymous referee.

in the `_initobs` equation.[28] To achieve this, technically we need to both drop the first-differenced variables and add the levels:

```
. xtdpdqml n w k yr1978-yr1984, stationary projection(w k, omit)
> projection(w k, nodifference) mlparams nolog
Quasi-maximum likelihood estimation

Group variable: id                         Number of obs        =       891
Time variable: year                        Number of groups     =       140

Fixed effects                              Obs per group:    min =         6
                                                             avg =  6.364286
                                                             max =         8

------------------------------------------------------------------------------
         D.n |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
_model       |
           n |
         LD. |   .7169499   .0348373    20.58   0.000       .64867    .7852298
             |
           w |
         D1. |  -.4231864   .0512345    -8.26   0.000    -.5236041   -.3227686
             |
           k |
         D1. |   .2501779   .0254106     9.85   0.000      .200374    .2999817
             |
      yr1978 |
         D1. |  -.0211017   .0149431    -1.41   0.158    -.0503896    .0081861
             |
      yr1979 |
         D1. |  -.0315607   .0149312    -2.11   0.035    -.0608253   -.0022961
             |
      yr1980 |
         D1. |  -.0628003   .0148639    -4.23   0.000    -.0919331   -.0336676
             |
      yr1981 |
         D1. |  -.1119848   .0150481    -7.44   0.000    -.1414784   -.0824911
             |
      yr1982 |
         D1. |  -.0832384    .016064    -5.18   0.000    -.1147233   -.0517536
             |
      yr1983 |
         D1. |   -.044769   .0196758    -2.28   0.023     -.083333   -.0062051
             |
      yr1984 |
         D1. |  -.0098343   .0240858    -0.41   0.683    -.0570416    .0373729
------------------------------------------------------------------------------
```

28. Phillips (2014) finds simulation evidence that using levels of the right-hand-side variables in (5) can lead to an improved performance over using first differences. This idea is similar to the use of lagged levels as instruments for the first-differenced equation as proposed by Anderson and Hsiao (1981, 1982) and Arellano and Bond (1991) in the context of instrumental-variables or GMM estimation.

```
_initobs     |
      yr1978 |
         D1. |    .056943   .0335663     1.70   0.090    -.0088458    .1227318
         FD. |   .0387561   .0200159     1.94   0.053    -.0004743    .0779865
             |
           w |
         L1. |  -.1708625   .0853048    -2.00   0.045    -.3380569   -.0036681
         --. |  -.3271109   .1564291    -2.09   0.037    -.6337063   -.0205155
         F1. |   .2625743   .1459992     1.80   0.072    -.0235788    .5487274
         F2. |   .0456837   .1010946     0.45   0.651    -.1524581    .2438255
         F3. |   .0217997   .1003578     0.22   0.828    -.1748979    .2184973
         F4. |  -.0289995   .0609608    -0.48   0.634    -.1484805    .0904816
         F5. |   .1955317   .0799724     2.44   0.014     .0387887    .3522746
             |
           k |
         L1. |  -.2545108   .0526105    -4.84   0.000    -.3576254   -.1513962
         --. |   .3322412   .0816502     4.07   0.000     .1722097    .4922727
         F1. |  -.1117364    .064621    -1.73   0.084    -.2383912    .0149184
         F2. |  -.0051431   .0526266    -0.10   0.922    -.1082892    .0980031
         F3. |   .0245103    .058062     0.42   0.673    -.0892891    .1383096
         F4. |  -.1203073   .0578168    -2.08   0.037    -.2336263   -.0069884
         F5. |   .1350255   .0419887     3.22   0.001     .0527292    .2173219
-------------+----------------------------------------------------------------
   /_sigma2e |   .0107329   .0005941                      .0095685    .0118972
     /_omega |   1.220817   .0689984                      1.085583    1.356051
------------------------------------------------------------------------------
```

The coefficients of main interest in the `_model` equation hardly differ from the earlier specification. Because first differencing is equivalent to imposing linear restrictions on the coefficients of the levels, we again have two nested models such that we can use another likelihood-ratio test to decide the preferred specification:

```
. lrtest fe_s

Likelihood-ratio test                                 LR chi2(2)  =      0.05
(Assumption: fe_s nested in .)                        Prob > chi2 =    0.9776
```

Clearly, there is no gain from using the levels of the exogenous variables instead of their first differences in the `_initobs` equation.

The above QML estimates all are based on “difference GMM” estimates as starting values for the iterative maximization algorithm. We can recover these initial estimates by specifying a name for them with the `storeinit()` option:[29]

29. The initial GMM estimates provide starting values for the parameters λ and $\boldsymbol{\beta}$ in (4). They are the same irrespective of the use of the `stationary` or `projection()` options, which only affect the parameters in (5). Starting values for the latter and for the variance parameters are obtained automatically by `xtdpdqml` based on the initial estimates of λ and $\boldsymbol{\beta}$; see the online appendix for the respective formula. The whole set of starting values can be displayed by combining the options `mlparams` and `iterate(0)`.

```
. xtdpdqml n w k yr1978-yr1984, stationary storeinit(gmm)
  (output omitted)
. estimates replay gmm
--------------------------------------------------------------------------------
Model gmm (initial estimates for xtdpdqml)
--------------------------------------------------------------------------------

Dynamic panel-data estimation                 Number of obs      =       891
Group variable: id                            Number of groups   =       140
Time variable: year
                                              Obs per group:
                                                           min =         6
                                                           avg =  6.364286
                                                           max =         8

Number of instruments =     38                Wald chi2(10)      =   2218.24
                                              Prob > chi2        =    0.0000
One-step results
------------------------------------------------------------------------------
           n |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           n |
         L1. |     .32667   .0768081     4.25   0.000     .1761289     .477211
             |
           w |  -.4763421   .0530543    -8.98   0.000    -.5803266   -.3723575
           k |   .3271291   .0280945    11.64   0.000     .2720648    .3821934
      yr1978 |  -.0285803   .0117587    -2.43   0.015    -.0516269   -.0055337
      yr1979 |  -.0359986   .0117658    -3.06   0.002    -.0590592   -.0129381
      yr1980 |  -.0637982   .0116966    -5.45   0.000    -.0867231   -.0408733
      yr1981 |   -.118767   .0120325    -9.87   0.000    -.1423503   -.0951837
      yr1982 |  -.1233297   .0159048    -7.75   0.000    -.1545025   -.0921569
      yr1983 |  -.1054798   .0225806    -4.67   0.000     -.149737   -.0612227
      yr1984 |  -.0878231   .0287666    -3.05   0.002    -.1442045   -.0314417
       _cons |   2.398147   .1740268    13.78   0.000     2.057061    2.739233
------------------------------------------------------------------------------
Instruments for differenced equation
        GMM-type: L(2/.).n
        Standard: D.w D.k D.yr1978 D.yr1979 D.yr1980 D.yr1981 D.yr1982
                  D.yr1983 D.yr1984
Instruments for level equation
        Standard: _cons
```

The initial estimates for the QML estimator can be overwritten with the `from()` option, for example, if a "system GMM" estimator is justified:[30]

```
. xtdpdsys n w k yr1978-yr1984, twostep
  (output omitted)
. matrix b = e(b)
. xtdpdqml n w k yr1978-yr1984, stationary from(b, skip)
  (output omitted)
. estimates store fe_eq1
```

30. Alternative starting values for the variance parameters σ_e^2 and ω could be supplied with the `initval()` option; see section 3.2. To be feasible, the starting value for ω needs to be larger than $(T-1)/T$, where $T = \max(T_i)$.

The QML estimator converges to the same results under both initializations. While the estimator of Hsiao, Pesaran, and Tahmiscioglu (2002) is consistent both in a fixed-effects and in a random-effects world, the Bhargava and Sargan (1983) estimator would be more efficient under random effects but inconsistent in the presence of fixed effects. In the present case, it turns out that the starting values for the variance parameters in the random-effects model are infeasible:

```
. xtdpdqml n w k yr1978-yr1984, re
Quasi-maximum likelihood estimation
initial values not feasible
r(1400);
```

We can supply alternative starting values for the variance parameters with the `initval()` option:[31]

```
. xtdpdqml n w k yr1978-yr1984, re initval(.1 .2 .2 .3) nolog
Quasi-maximum likelihood estimation
Group variable: id                          Number of obs        =      1031
Time variable: year                         Number of groups     =       140
Random effects                              Obs per group:   min =         7
                                                             avg =  7.364286
                                                             max =         9
```

n	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
n						
L1.	.6827452	.0264103	25.85	0.000	.6309818	.7345085
w	-.3044987	.0422166	-7.21	0.000	-.3872419	-.2217556
k	.2630637	.0214881	12.24	0.000	.2209478	.3051796
yr1978	-.0215183	.0148306	-1.45	0.147	-.0505856	.0075491
yr1979	-.0326742	.0148093	-2.21	0.027	-.0616998	-.0036485
yr1980	-.0639498	.014763	-4.33	0.000	-.0928847	-.0350148
yr1981	-.1171753	.0148591	-7.89	0.000	-.1462986	-.0880519
yr1982	-.0953543	.0151577	-6.29	0.000	-.1250629	-.0656457
yr1983	-.0651054	.0180881	-3.60	0.000	-.1005575	-.0296533
yr1984	-.035986	.0226091	-1.59	0.111	-.0802991	.0083271
_cons	1.437169	.1517993	9.47	0.000	1.139648	1.73469

```
. estimates store re_eq1
```

For the random-effects QML estimator, a specification test rejects the restriction imposed by the stationarity assumptions on the dependent variable and the regressors:[32]

---

31. The `initval()` option specifies starting values for the variance parameters in the order σ_u^2, σ_e^2, σ_0^2, and ϕ; see section 3.2. The chosen values should satisfy the particular constraint $(\sigma_u^2 - \phi^2\sigma_0^2)T > -\sigma_e^2$, where $T = \max(T_i)$, taking into account the restrictions on the variance parameters if the `stationary` option is specified; see section 2.1 and the online appendix for details.
32. In contrast to the fixed-effects model, the restriction in the random-effects model is not on the initial-observations intercept but on the covariance between the initial-observations error term and the unit-specific effects. Moreover, in the fixed-effects model, first-difference stationarity of the regressors is required compared with stationarity of the levels in the random-effects model, aside from $|\lambda| < 1$ and an initialization of the process in the very past; see section 2.

```
. xtdpdqml n w k yr1978-yr1984, re stationary initval(.1 .2 .2 .3)
  (output omitted)
. lrtest re_eq1
Likelihood-ratio test                                 LR chi2(1)  =      4.05
(Assumption: . nested in re_eq1)                      Prob > chi2 =    0.0443
```

Assuming that the model is correctly specified, we can employ the traditional Hausman (1978) test to discern between the fixed-effects and the random-effects models:[33]

```
. hausman fe_eq1 re_eq1, df(3)
                 ---- Coefficients ----
             |      (b)          (B)            (b-B)     sqrt(diag(V_b-V_B))
             |     fe_eq1       re_eq1       Difference          S.E.
-------------+----------------------------------------------------------------
           n |
         L1. |    .7175701     .6827452         .034825        .0226023
           w |   -.4219682    -.3044987       -.1174694        .0284716
           k |    .2493912     .2630637       -.0136726        .0131215
      yr1978 |   -.0212959    -.0215183        .0002224        .0016011
      yr1979 |   -.0317929    -.0326742        .0008813        .0015725
      yr1980 |   -.0633101    -.0639498        .0006397               .
      yr1981 |   -.1125881    -.1171753        .0045871               .
      yr1982 |   -.0839164    -.0953543        .0114378        .0042314
      yr1983 |   -.0455604    -.0651054         .019545         .006765
      yr1984 |   -.0107753     -.035986        .0252107        .0069979
------------------------------------------------------------------------------
                          b = consistent under Ho and Ha; obtained from xtdpdqml
           B = inconsistent under Ha, efficient under Ho; obtained from xtdpdqml

    Test:  Ho:  difference in coefficients not systematic

                  chi2(3) = (b-B)'[(V_b-V_B)^(-1)](b-B)
                          =      239.87
                Prob>chi2 =      0.0000
                (V_b-V_B is not positive definite)
```

The null hypothesis is strongly rejected in favor of the fixed-effects model. Notice that the **hausman** command was executed with the option **df(3)**. The reason is that we cannot include the time effects in the comparison because of a singularity in the asymptotic covariance matrix of the difference between the fixed-effects and the random-effects estimates. The degrees of freedom therefore equals the number of time-varying regressors excluding the time dummies.[34]

33. The `hausman` command cannot compare the fixed-effects coefficients reported for the first-differenced equation with the random-effects coefficients for the levels equation. Therefore, we had to estimate the models without the `mlparams` option such that all coefficients are reported for the levels equation. The remaining parameters are not needed for this test.
34. See Wooldridge (2010, chap. 10.7.3) for a discussion of this problem.

An underlying assumption of this Hausman test is that one of the estimators is efficient. However, this would no longer be the case if the model is misspecified. For example, the assumed error covariance structure might be invalid. Exemplary for the fixed-effects estimator, we observe that the standard errors indeed increase sizably if we use a variance–covariance estimator that is robust to cross-sectional heteroskedasticity with the option `vce(robust)`:[35]

```
. xtdpdqml n w k yr1978-yr1984, stationary vce(robust) nolog
Quasi-maximum likelihood estimation

Group variable: id                           Number of obs        =       891
Time variable: year                          Number of groups     =       140

Fixed effects                                Obs per group:    min =         6
                                                               avg =  6.364286
(Estimation in first differences)                              max =         8
                                       (Std. Err. adjusted for clustering on id)

------------------------------------------------------------------------------
             |               Robust
           n |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           n |
         L1. |   .7175702   .0777459     9.23   0.000     .5651911    .8699493
             |
           w |  -.4219682   .1284262    -3.29   0.001    -.6736789   -.1702574
           k |   .2493911   .0459749     5.42   0.000      .159282    .3395002
      yr1978 |  -.0212959   .0140127    -1.52   0.129    -.0487603    .0061685
      yr1979 |  -.0317929   .0163543    -1.94   0.052    -.0638468     .000261
      yr1980 |  -.0633101   .0172692    -3.67   0.000    -.0971571   -.0294631
      yr1981 |  -.1125881   .0196329    -5.73   0.000     -.151068   -.0741083
      yr1982 |  -.0839164   .0174937    -4.80   0.000    -.1182034   -.0496294
      yr1983 |  -.0455604   .0201398    -2.26   0.024    -.0850337   -.0060871
      yr1984 |  -.0107753   .0265246    -0.41   0.685    -.0627626     .041212
       _cons |   1.751776   .4532361     3.87   0.000     .8634493    2.640102
------------------------------------------------------------------------------
```

If there is cross-sectional heteroskedasticity, the results from the traditional Hausman test are no longer valid because the random-effects QML estimator is not efficient any more, and the `hausman` command refuses to accept estimates with a robust variance–covariance estimator. As a feasible alternative, we can use Stata's `suest` command to perform a generalized Hausman test. `suest` can combine the fixed-effects and the random-effects QML estimates, and it calculates a simultaneous variance–covariance estimator. For the latter, `suest` needs the nonrobust variance–covariance estimate for the whole set of parameters as input. Thus, we are requested to execute `xtdpdqml` without the option `vce(robust)` but with the `mlparams` option.[36] Moreover, we should specify the option `vce(cluster id)` when calling `suest` to account for the panel structure:

35. `vce(robust)` causes the sandwich formula to be used for the variance–covariance estimator, taking into account that the observations are not independent across time. Robust standard errors are also calculated for the initial GMM estimates if they are stored with the `storeinit()` option. Neither the initial values nor the final coefficient estimates are affected by the type of standard errors. Hayakawa and Pesaran (2015) demonstrate that the QML estimator remains consistent under cross-sectional heteroskedasticity. However, it becomes inconsistent if the untransformed idiosyncratic error component is serially correlated or heteroskedastic across time.

36. `suest` also requires the postestimation command `predict` to produce equation-level scores. After `xtdpdqml`, this is possible only with the `mlparams` option and not with the `stationary` option.

```
. xtdpdqml n w k yr1978-yr1984, re initval(.1 .2 .2 .3) mlparams
  (output omitted)
. estimates store re
. suest fe re, vce(cluster id)
  (output omitted)
. test ([fe__model]LD.n = [re__model]L.n) ([fe__model]D.w = [re__model]w)
> ([fe__model]D.k = [re__model]k)
 ( 1)  [fe__model]LD.n - [re__model]L.n = 0
 ( 2)  [fe__model]D.w - [re__model]w = 0
 ( 3)  [fe__model]D.k - [re__model]k = 0
           chi2(  3) =    5.97
         Prob > chi2 =    0.1132
```

The `suest` output is omitted because of its length. It simultaneously reports the results for both models as separate equations. Further above, we saw the output generated by `xtdpdqml` with the `mlparams` option. It separates the parameters into different equations, with the first two named `_model` and `_initobs` for our model parameters of main interest and the initial-observations coefficients, respectively. `suest` builds its equation names as combinations of the estimates' name and `xtdpdqml`'s equation names, resulting in `fe__model`, `fe__initobs`, `re__model`, `re__initobs`, and similarly for the variance parameters that make up separate equations. This helps us to understand the syntax of the `test` command after `suest`. To test for systematic differences between the fixed-effects and the random-effects models, we need to compare the coefficients of the lagged dependent variable and the two exogenous regressors `w` and `k` from the two models.[37] As we can see, the Wald test does not reject the null hypothesis at the conventional significance levels, which is the opposite result from the traditional Hausman test above. However, we should not be too confident about this result because the p-value is still relatively small. With the fixed-effects QML estimator, we remain on the safe side.

Finally, after fitting our model, we might be interested in the long-run effects of the exogenous regressors. While the coefficients $\boldsymbol{\beta}$ are short-run effects conditional on the initial level of employment, `L.n`, the corresponding long-run effects can be computed as $\boldsymbol{\beta}/(1-\lambda)$. In Stata, we can obtain such estimates with the `nlcom` command:

```
. xtdpdqml n w k yr1978-yr1984, stationary vce(robust)
  (output omitted)
. nlcom (_b[w] / (1 - _b[L.n])) (_b[k] / (1 - _b[L.n]))
       _nl_1:  _b[w] / (1 - _b[L.n])
       _nl_2:  _b[k] / (1 - _b[L.n])
```

n	Coef.	Std. Err.	z	P>|z|	[95% Conf. Interval]	
_nl_1	-1.494064	.4484327	-3.33	0.001	-2.372976	-.6151519
_nl_2	.8830199	.1834742	4.81	0.000	.523417	1.242623

37. With the same reasoning as before, the time dummies are not included in the comparison.

6 Conclusion

In this article, I presented the new estimation command `xtdpdqml`, which extends the available toolkit for linear dynamic panel model estimation in Stata. It implements QML estimators for random-effects and fixed-effects models that account for the endogeneity of the initial observations to avoid biased estimates when the time dimension is short.

7 Acknowledgments

I thank Michael Binder for his encouragement to work on this topic, and Georgios Georgiadis and Daniel Schneider for valuable input. Comments and suggestions by Paul Allison, Alok Bhargava, Cheng Hsiao, Hashem Pesaran, and one anonymous referee are highly appreciated, as well as the provision of MATLAB code for comparison by Kazuhiko Hayakawa.

8 References

Anderson, T. W., and C. Hsiao. 1981. Estimation of dynamic models with error components. *Journal of the American Statistical Association* 76: 598–606.

———. 1982. Formulation and estimation of dynamic models using panel data. *Journal of Econometrics* 18: 47–82.

Arellano, M., and S. Bond. 1991. Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. *Review of Economic Studies* 58: 277–297.

Arellano, M., and O. Bover. 1995. Another look at the instrumental variable estimation of error-components models. *Journal of Econometrics* 68: 29–51.

Bhargava, A., and J. D. Sargan. 1983. Estimating dynamic random effects models from panel data covering short time periods. *Econometrica* 51: 1635–1659.

Blundell, R., and S. Bond. 1998. Initial conditions and moment restrictions in dynamic panel data models. *Journal of Econometrics* 87: 115–143.

Bontempi, M. E., and I. Mammi. 2015. Implementing a strategy to reduce the instrument count in panel GMM. *Stata Journal* 15: 1075–1097.

Bruno, G. S. F. 2005. Estimation and inference in dynamic unbalanced panel-data models with a small number of individuals. *Stata Journal* 5: 473–500.

Bun, M. J. G., and J. F. Kiviet. 2003. On the diminishing returns of higher-order terms in asymptotic expansions of bias. *Economics Letters* 79: 145–152.

———. 2006. The effects of dynamic feedbacks on LS and MM estimator accuracy in panel data models. *Journal of Econometrics* 132: 409–444.

De Vos, I., G. Everaert, and I. Ruyssen. 2015. Bootstrap-based bias correction and inference for dynamic panels with fixed effects. *Stata Journal* 15: 986–1018.

Everaert, G., and L. Pozzi. 2007. Bootstrap-based bias correction for dynamic panels. *Journal of Economic Dynamics and Control* 31: 1160–1184.

Gould, W., J. Pitblado, and B. Poi. 2010. *Maximum Likelihood Estimation with Stata*. 4th ed. College Station, TX: Stata Press.

Hausman, J. A. 1978. Specification tests in econometrics. *Econometrica* 46: 1251–1271.

Hayakawa, K., and M. H. Pesaran. 2015. Robust standard errors in transformed likelihood estimation of dynamic panel data models with cross-sectional heteroskedasticity. *Journal of Econometrics* 188: 111–134.

Hsiao, C., M. H. Pesaran, and A. K. Tahmiscioglu. 2002. Maximum likelihood estimation of fixed effects dynamic panel data models covering short time periods. *Journal of Econometrics* 109: 107–150.

Kiviet, J. F. 1995. On bias, inconsistency, and efficiency of various estimators in dynamic panel data models. *Journal of Econometrics* 68: 53–78.

Kripfganz, S., and C. Schwarz. 2015. Estimation of linear dynamic panel data models with time-invariant regressors. ECB Working Paper 1838, European Central Bank.

Nickell, S. 1981. Biases in dynamic models with fixed effects. *Econometrica* 49: 1417–1426.

Phillips, R. F. 2014. Quasi maximum-likelihood estimation of dynamic panel data models for short time series. RPF Working Paper 2014-006, Research Program on Forecasting, The George Washington University.

Roodman, D. 2009. How to do xtabond2: An introduction to difference and system GMM in Stata. *Stata Journal* 9: 86–136.

Williams, R., P. Allison, and E. Moral-Benito. 2015. Linear dynamic panel-data estimation using maximum likelihood and structural equation modeling. Presented July 30, 2015 at the Stata Conference 2015, Columbus, Ohio.

Wooldridge, J. M. 2010. *Econometric Analysis of Cross Section and Panel Data*. 2nd ed. Cambridge, MA: MIT Press.

About the author

Sebastian Kripfganz joined the University of Exeter Business School in September 2015 as a lecturer in economics after finishing his doctorate at Goethe University Frankfurt. Dynamic panel-data econometrics ranks prominently among his research interests.