Dynamic Panel Data estimators

Christopher F Baum

EC 823: Applied Econometrics

Boston College, Spring 2014

Dynamic panel data estimators

In the context of panel data, we usually must deal with unobserved heterogeneity by applying the within (demeaning) transformation, as in one-way fixed effects models, or by taking first differences if the second dimension of the panel is a proper time series.

The ability of first differencing to remove unobserved heterogeneity also underlies the family of estimators that have been developed for dynamic panel data (DPD) models. These models contain one or more lagged dependent variables, allowing for the modeling of a partial adjustment mechanism.

Nickell bias

A serious difficulty arises with the one-way fixed effects model in the context of a *dynamic panel data* (DPD) model particularly in the “small T, large N" context. As Nickell (*Econometrica*, 1981) shows, this arises because the demeaning process which subtracts the individual’s mean value of y and each X from the respective variable creates a correlation between regressor and error.

The mean of the lagged dependent variable contains observations 0 through $(T-1)$ on y, and the mean error—which is being conceptually subtracted from each ϵ_{it}—contains contemporaneous values of ϵ for $t = 1 \ldots T$. The resulting correlation creates a bias in the estimate of the coefficient of the lagged dependent variable which is not mitigated by increasing N, the number of individual units.

The demeaning operation creates a regressor which *cannot* be distributed independently of the error term. Nickell demonstrates that the inconsistency of $\hat{\rho}$ as $N \rightarrow \infty$ is of order $1/T$, which may be quite sizable in a “small T" context. If $\rho > 0$, the bias is invariably negative, so that the persistence of y will be underestimated.

For reasonably large values of T, the limit of $(\hat{\rho} - \rho)$ as $N \rightarrow \infty$ will be approximately $-(1 + \rho)/(T - 1)$: a sizable value, even if $T = 10$. With $\rho = 0.5$, the bias will be -0.167, or about 1/3 of the true value. The inclusion of additional regressors does not remove this bias. Indeed, if the regressors are correlated with the lagged dependent variable to some degree, their coefficients may be seriously biased as well.

Note also that this bias is not caused by an autocorrelated error process ϵ. The bias arises even if the error process is *i.i.d*. If the error process is autocorrelated, the problem is even more severe given the difficulty of deriving a consistent estimate of the *AR* parameters in that context.

The same problem affects the one-way random effects model. The u_i error component enters every value of y_{it} by assumption, so that the lagged dependent variable *cannot* be independent of the composite error process.

One solution to this problem involves taking first differences of the original model. Consider a model containing a lagged dependent variable and a single regressor X:

$$y_{it} = \beta_1 + \rho y_{i,t-1} + X_{it}\beta_2 + \; u_i + \epsilon_{it} \tag{1}$$

The first difference transformation removes both the constant term and the individual effect:

$$\Delta y_{it} = \rho \Delta y_{i,t-1} + \Delta X_{it}\beta_2 + \Delta \epsilon_{it} \tag{2}$$

There is still correlation between the differenced lagged dependent variable and the disturbance process (which is now a first-order moving average process, or $MA(1)$): the former contains $y_{i,t-1}$ and the latter contains $\epsilon_{i,t-1}$.

But with the individual fixed effects swept out, a straightforward instrumental variables estimator is available. We may construct instruments for the lagged dependent variable from the second and third lags of y, either in the form of differences or lagged levels. If ϵ is $i.i.d.$, those lags of y will be highly correlated with the lagged dependent variable (and its difference) but uncorrelated with the composite error process.

Even if we had reason to believe that ϵ might be following an $AR(1)$ process, we could still follow this strategy, "backing off" one period and using the third and fourth lags of y (presuming that the timeseries for each unit is long enough to do so).

This approach is the Anderson–Hsiao (AH) estimator implemented by the Stata command `xtivreg, fd`.

The DPD approach

The *DPD* (Dynamic Panel Data) approach is usually considered the work of Arellano and Bond (AB) (*Rev. Ec. Stud.*, 1991), but they in fact popularized the work of Holtz-Eakin, Newey and Rosen (*Econometrica*, 1988). It is based on the notion that the instrumental variables approach noted above does not exploit all of the information available in the sample. By doing so in a Generalized Method of Moments (GMM) context, we may construct more efficient estimates of the dynamic panel data model.

Arellano and Bond argue that the Anderson–Hsiao estimator, while consistent, fails to take all of the potential orthogonality conditions into account. A key aspect of the AB strategy, echoing that of AH, is the assumption that the necessary instruments are ‘internal’: that is, based on lagged values of the instrumented variable(s). The estimators allow the inclusion of external instruments as well.

Consider the equations

$$\begin{aligned} y_{it} &= X_{it}\beta_1 + W_{it}\beta_2 + v_{it} \\ v_{it} &= u_i + \epsilon_{it} \end{aligned} \tag{3}$$

where X_{it} includes strictly exogenous regressors, W_{it} are predetermined regressors (which may include lags of y) and endogenous regressors, all of which may be correlated with u_i, the unobserved individual effect. First-differencing the equation removes the u_i and its associated omitted-variable bias.

The AB approach, and its extension to the 'System GMM' context, is an estimator designed for situations with:

- 'small T, large N' panels: few time periods and many individual units
- a linear functional relationship
- one left-hand variable that is dynamic, depending on its own past realisations
- right-hand variables that are not strictly exogenous: correlated with past and possibly current realisations of the error
- fixed individual effects, implying unobserved heterogeneity
- heteroskedasticity and autocorrelation within individual units' errors, but not across them

The Arellano–Bond estimator sets up a generalized method of moments (*GMM*) problem in which the model is specified as a system of equations, one per time period, where the instruments applicable to each equation differ (for instance, in later time periods, additional lagged values of the instruments are available).

This estimator is available in Stata as `xtabond`. A more general version, allowing for autocorrelated errors, is available as `xtdpd`. An excellent alternative to Stata's built-in commands is David Roodman's `xtabond2`, available from SSC (`findit xtabond2`). It is very well documented in his paper, included in your materials. The `xtabond2` routine provides several additional features—such as the orthogonal deviations transformation discussed below—not available in official Stata's commands.

Constructing the instrument matrix

In standard 2SLS, including the Anderson–Hsiao approach, the twice-lagged level appears in the instrument matrix as

$$\mathbf{Z}_i = \begin{pmatrix} . \\ y_{i,1} \\ \vdots \\ y_{i,T-2} \end{pmatrix}$$

where the first row corresponds to $t = 2$, given that the first observation is lost in applying the FD transformation. The missing value in the instrument for $t = 2$ causes that observation for each panel unit to be removed from the estimation.

If we also included the thrice-lagged level y_{t-3} as a second instrument in the Anderson–Hsiao approach, we would lose another observation per panel:

$$\mathbf{Z}_i = \begin{pmatrix} . & . \\ y_{i,1} & . \\ y_{i,2} & y_{i,1} \\ \vdots & \vdots \\ y_{i,T-2} & y_{i,T-3} \end{pmatrix}$$

so that the first observation available for the regression is that dated $t = 4$.

To avoid this loss of degrees of freedom, Holtz-Eakin et al. construct a set of instruments from the second lag of y, one instrument pertaining to each time period:

$$\mathbf{Z}_i = \begin{pmatrix} 0 & 0 & \dots & 0 \\ y_{i,1} & 0 & \dots & 0 \\ 0 & y_{i,2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & y_{i,T-2} \end{pmatrix}$$

The inclusion of zeros in place of missing values prevents the loss of additional degrees of freedom, in that all observations dated $t = 2$ and later can now be included in the regression. Although the inclusion of zeros might seem arbitrary, the columns of the resulting instrument matrix will be orthogonal to the transformed errors. The resulting moment conditions correspond to an expectation we believe should hold: $E(y_{i,t-2}\epsilon^*_{it}) = 0$, where ϵ^* refers to the FD-transformed errors.

It would also be valid to 'collapse' the columns of this **Z** matrix into a single column, which embodies the same expectation, but conveys less information as it will only produce a single moment condition. In this context, the collapsed instrument set will be the same implied by standard IV, with a zero replacing the missing value in the first usable observation:

$$\mathbf{Z}_i = \begin{pmatrix} 0 \\ y_{i,1} \\ \vdots \\ y_{i,T-2} \end{pmatrix}$$

This is specified in Roodman's `xtabond2` software by giving the `collapse` option.

Given this solution to the tradeoff between lag length and sample length, we can now adopt Holtz-Eakin et al.'s suggestion and include *all* available lags of the untransformed variables as instruments. For endogenous variables, lags 2 and higher are available. For predetermined variables that are not strictly exogenous, lag 1 is also valid, as its value is only correlated with errors dated $t-2$ or earlier.

Using all available instruments gives rise to an instrument matrix such as

$$\mathbf{Z}_i = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & \ldots \\ y_{i,1} & 0 & 0 & 0 & 0 & 0 & \ldots \\ 0 & y_{i,2} & y_{i,1} & 0 & 0 & 0 & \ldots \\ 0 & 0 & 0 & y_{i,3} & y_{i,2} & y_{i,1} & \ldots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}$$

In this setup, we have different numbers of instruments available for each time period: one for $t = 2$, two for $t = 3$, and so on. As we move to the later time periods in each panel's timeseries, additional orthogonality conditions become available, and taking these additional conditions into account improves the efficiency of the AB estimator.

One disadvantage of this strategy should be apparent. The number of instruments produced will be quadratic in T, the length of the timeseries available. If $T < 10$, that may be a manageable number, but for a longer timeseries, it may be necessary to restrict the number of past lags used. Both the official Stata commands and Roodman's `xtabond2` allow the specification of the particular lags to be included in estimation, rather than relying on the default strategy.

The System GMM estimator

A potential weakness in the Arellano–Bond *DPD* estimator was revealed in later work by Arellano and Bover (1995) and Blundell and Bond (1998). The lagged levels are often rather poor instruments for first differenced variables, especially if the variables are close to a random walk. Their modification of the estimator includes lagged levels as well as lagged differences.

The original estimator is often entitled *difference GMM*, while the expanded estimator is commonly termed *System GMM*. The cost of the System GMM estimator involves a set of additional restrictions on the initial conditions of the process generating y. This estimator is available in Stata as `xtdpdsys`.

Diagnostic tests

As the DPD estimators are instrumental variables methods, it is particularly important to evaluate the Sargan–Hansen test results when they are applied. Roodman's `xtabond2` provides *C* tests (as discussed in re `ivreg2`) for groups of instruments. In his routine, instruments can be either "GMM-style" or "IV-style". The former are constructed per the Arellano–Bond logic, making use of multiple lags; the latter are included as is in the instrument matrix. For the system GMM estimator (the default in `xtabond2`) instruments may be specified as applying to the differenced equations, the level equations or both.

Another important diagnostic in DPD estimation is the *AR* test for autocorrelation of the residuals. By construction, the residuals of the differenced equation should possess serial correlation, but if the assumption of serial independence in the original errors is warranted, the differenced residuals should not exhibit significant *AR*(2) behavior. These statistics are produced in the `xtabond` and `xtabond2` output. If a significant *AR*(2) statistic is encountered, the second lags of endogenous variables will not be appropriate instruments for their current values.

A useful feature of `xtabond2` is the ability to specify, for GMM-style instruments, the limits on how many lags are to be included. If T is fairly large (more than 7–8) an unrestricted set of lags will introduce a huge number of instruments, with a possible loss of efficiency. By using the lag limits options, you may specify, for instance, that only lags 2–5 are to be used in constructing the GMM instruments.

An empirical exercise

To illustrate the performance of the several estimators, we make use of the original AB dataset, available within Stata with `webuse abdata`. This is an unbalanced panel of annual data from 140 UK firms for 1976–1984. In their original paper, they modeled firms' employment `n` using a partial adjustment model to reflect the costs of hiring and firing, with two lags of employment.

Other variables included were the current and lagged wage level `w`, the current, once- and twice-lagged capital stock (`k`) and the current, once- and twice-lagged output in the firm's sector (`ys`). All variables are expressed as logarithms. A set of time dummies is also included to capture business cycle effects.

If we were to estimate this model ignoring its dynamic panel nature, we could merely apply `regress` with panel-clustered standard errors:

```
regress n nL1 nL2 w wL1 k kL1 kL2 ys ysL1 ysL2 yr*, cluster(id)
```

One obvious difficulty with this approach is the likely importance of firm-level unobserved heterogeneity. We have accounted for potential correlation between firms' errors over time with the cluster-robust VCE, but this does not address the potential impact of unobserved heterogeneity on the conditional mean.

We can apply the within transformation to take account of this aspect of the data:

```
xtreg n nL1 nL2 w wL1 k kL1 kL2 ys ysL1 ysL2 yr*, fe cluster(id)
```

The fixed effects estimates will suffer from Nickell bias, which may be severe given the short timeseries available.

	OLS		FE	
nL1	1.045^{***}	(20.17)	0.733^{***}	(12.28)
nL2	-0.0765	(-1.57)	-0.139	(-1.78)
w	-0.524^{**}	(-3.01)	-0.560^{***}	(-3.51)
k	0.343^{***}	(7.06)	0.388^{***}	(6.82)
ys	0.433^{*}	(2.42)	0.469^{**}	(2.74)
N	751		751	

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

In the original OLS regression, the lagged dependent variable was positively correlated with the error, biasing its coefficient upward. In the fixed effects regression, its coefficient is biased downward due to the negative sign on ν_{t-1} in the transformed error. The OLS estimate of the first lag of *n* is 1.045; the fixed effects estimate is 0.733.

Given the opposite directions of bias present in these estimates, consistent estimates should lie between these values, which may be a useful check. As the coefficient on the second lag of *n* cannot be distinguished from zero, the first lag coefficient should be below unity for dynamic stability.

To deal with these two aspects of the estimation problem, we might use the Anderson–Hsiao estimator to the first-differenced equation, instrumenting the lagged dependent variable with the twice-lagged level:

```
 ivregress 2sls D.n (D.nL1 = nL2) D.(nL2 w wL1 k kL1 kL2 ///
ys ysL1 ysL2 yr1979 yr1980 yr1981 yr1982 yr1983 )
```

	A-H	
D.nL1	2.308	(1.17)
D.nL2	-0.224	(-1.25)
D.w	-0.810**	(-3.10)
D.k	0.253	(1.75)
D.ys	0.991*	(2.14)
N	611	

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Although these results should be consistent, they are quite disappointing. The coefficient on lagged *n* is outside the bounds of its OLS and FE counterparts, and much larger than unity, a value consistent with dynamic stability. It is also very imprecisely estimated.

The difference GMM approach deals with this inherent endogeneity by transforming the data to remove the fixed effects. The standard approach applies the first difference (FD) transformation, which as discussed earlier removes the fixed effect at the cost of introducing a correlation between $\Delta y_{i,t-1}$ and $\Delta \nu_{it}$, both of which have a term dated $(t-1)$. This is preferable to the application of the within transformation, as that transformation makes every observation in the transformed data endogenous to every other for a given individual.

The one disadvantage of the first difference transformation is that it magnifies gaps in unbalanced panels. If some value of y_{it} is missing, then both Δy_{it} and $\Delta y_{i,t-1}$ will be missing in the transformed data. This motivates an alternative transformation: the forward orthogonal deviations (FOD) transformation, proposed by Arellano and Bover (*J. Econometrics*, 1995).

In contrast to the within transformation, which subtracts the average of all observations' values from the current value, and the FD transformation, that subtracts the previous value from the current value, the FOD transformation subtracts the average of all available *future* observations from the current value. While the FD transformation drops the first observation on each individual in the panel, the FOD transformation drops the last observation for each individual. It is computable for all periods except the last period, even in the presence of gaps in the panel.

The FOD transformation is not available in any of official Stata's DPD commands, but it is available in David Roodman's `xtabond2` implementation of the DPD estimator, available from SSC.

To illustrate the use of the AB estimator, we may reestimate the model with `xtabond2`, assuming that the only endogeneity present is that involving the lagged dependent variable.

```
 xtabond2 n L(1/2).n L(0/1).w L(0/2).(k ys) yr*, gmm(L.n) ///
iv(L(0/1).w L(0/2).(k ys) yr*) nolevel robust small
```

Note that in `xtabond2` syntax, every right-hand variable generally appears twice in the command, as instruments must be explicitly specified when they are instrumenting themselves. In this example, all explanatory variables except the lagged dependent variable are taken as “IV-style” instruments, entering the **Z** matrix as a single column. The lagged dependent variable is specified as a “GMM-style” instrument, where all available lags will be used as separate instruments. The `noleveleq` option is needed to specify the AB estimator.

	A-B	
L.n	0.686^{***}	(4.67)
L2.n	-0.0854	(-1.50)
w	-0.608^{**}	(-3.36)
k	0.357^{***}	(5.95)
ys	0.609^{***}	(3.47)
N	611	

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

In these results, 41 instruments have been created, with 17 corresponding to the "IV-style" regressors and the rest computed from lagged values of n. Note that the coefficient on the lagged dependent variable now lies within the range for dynamic stability. In contrast to that produced by the Anderson–Hsiao estimator, the coefficient is quite precisely estimated.

There are 25 overidentifying restrictions in this instance, as shown in the first column below. The `hansen_df` represents the degrees of freedom for the Hansen *J* test of overidentifying restrictions. The *p*-value of that test is shown as `hansenp`.

	All lags		lags 2-5		lags 2-4	
L.n	0.686***	(4.67)	0.835*	(2.59)	1.107***	(3.94)
L2.n	-0.0854	(-1.50)	0.262	(1.56)	0.231	(1.32)
w	-0.608**	(-3.36)	-0.671**	(-3.18)	-0.709**	(-3.26)
k	0.357***	(5.95)	0.325***	(4.95)	0.309***	(4.55)
ys	0.609***	(3.47)	0.640**	(3.07)	0.698***	(3.45)
hansen_df	25		16		13	
hansenp	0.177		0.676		0.714	

t statistics in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

In this table, we can examine the sensitivity of the results to the choice of “GMM-style” lag specification. In the first column, all available lags of the level of *n* are used. In the second column, the `lag(2 5)` option is used to restrict the maximum lag to 5 periods, while in the third column, the maximum lag is set to 4 periods. Fewer instruments are used in those instances, as shown by the smaller values of `sar_df`.

The *p*-value of Hansen’s *J* is also considerably larger for the restricted-lag cases. On the other hand, the estimate of the lagged dependent variable’s coefficient appears to be quite sensitive to the choice of lag length.

We illustrate estimating this equation with both the FD transformation and the forward orthogonal deviations (FOD) transformation:

	First diff		FOD	
L.n	0.686***	(4.67)	0.737***	(5.14)
L2.n	-0.0854	(-1.50)	-0.0960	(-1.38)
w	-0.608**	(-3.36)	-0.563***	(-3.47)
k	0.357***	(5.95)	0.384***	(6.85)
ys	0.609***	(3.47)	0.469**	(2.72)
hansen_df	25		25	
hansenp	0.177		0.170	

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

The results appear reasonably robust to the choice of transformation, with slightly more precise estimates for most coefficients when the FOD transformation is employed.

We might reasonably consider, as did Blundell and Bond (*J. Econometrics*, 1998), that wages and the capital stock should not be taken as strictly exogenous in this context, as we have in the above models.

Reestimate the equation producing “GMM-style” instruments for all three variables, with both one-step and two-step VCE:

```
 xtabond2 n L(1/2).n L(0/1).w L(0/2).(k ys) yr*, gmm(L.(n w k)) ///
iv(L(0/2).ys yr*) nolevel robust small
```

	One-step		Two-step	
L.n	0.818***	(9.51)	0.824***	(8.51)
L2.n	-0.112*	(-2.23)	-0.101	(-1.90)
w	-0.682***	(-4.78)	-0.711***	(-4.67)
k	0.353**	(2.89)	0.377**	(2.79)
ys	0.651***	(3.43)	0.662***	(3.89)
hansen_df	74		74	
hansenp	0.487		0.487	

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

The results from both one-step and two-step estimation appear reasonable. Interestingly, only the coefficient on *ys* appears to be more precisely estimated by the two-step VCE. With no restrictions on the instrument set, 74 overidentifying restrictions are defined, with 90 instruments in total.

To illustrate system GMM, we follow Blundell and Bond, who used the same `abdata` dataset on a somewhat simpler model, dropping the second lags and removing sectoral demand. We consider wages and capital as potentially endogenous, with GMM-style instruments.

Estimate the one-step BB model.

```
xtabond2 n L.n L(0/1).(w k) yr*, gmm(L.(n w k)) iv(yr*, equation(level)) ///
robust small
```

We indicate here with the `equation(level)` suboption that the year dummies are only to be considered instruments in the level equation. As the default for `xtabond2` is the BB estimator, we omit the `noleveleq` option that has called for the AB estimator in earlier examples.

	n	
L.n	0.936***	(35.21)
w	-0.631***	(-5.29)
k	0.484***	(8.89)
hansen_df	100	
hansenp	0.218	

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

We find that the α coefficient is much higher than in the AB estimates, although it may be distinguished from unity. 113 instruments are created, with 100 degrees of freedom in the test of overidentifying restrictions.

A second empirical exercise

We also illustrate DPD estimation using the Penn World Table cross-country panel. We specify a model for `kc` (the consumption share of real GDP per capita) depending on its own lag, `cgnp`, and a set of time fixed effects, which we compute with the `xi` command, as `xtabond2` does not support factor variables.

We first estimate the two-step ‘difference GMM’ form of the model with (cluster-)robust VCE, using data for 1991–2007. We could use `testparm _I*` after estimation to evaluate the joint significance of time effects (listing of which has been suppressed).

```
. xi i.year
i.year            _Iyear_1991-2007    (naturally coded; _Iyear_1991 omitted)

. xtabond2 kc L.kc cgnp _I*, gmm(L.kc openc cgnp, lag(2 9)) iv(_I*) ///
> twostep robust noleveleq nodiffsargan
Favoring speed over space. To switch, type or click on mata: mata set matafavor
>  space, perm.

Dynamic panel-data estimation, two-step difference GMM
------------------------------------------------------------------------------
Group variable: iso                             Number of obs      =      1485
Time variable : year                            Number of groups   =        99
Number of instruments = 283                     Obs per group: min =        15
Wald chi2(17) =     94.96                                      avg =     15.00
Prob > chi2   =     0.000                                      max =        15
------------------------------------------------------------------------------
             |              Corrected
          kc |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          kc |
         L1. |   .6478636   .1041122     6.22   0.000     .4438075    .8519197
             |
        cgnp |    .233404   .1080771     2.16   0.031     .0215768    .4452312
...
------------------------------------------------------------------------------
```

(continued)

```
------------------------------------------------------------------------------
Instruments for first differences equation
  Standard
    D.(_Iyear_1992 _Iyear_1993 _Iyear_1994 _Iyear_1995 _Iyear_1996 _Iyear_1997
    _Iyear_1998 _Iyear_1999 _Iyear_2000 _Iyear_2001 _Iyear_2002 _Iyear_2003
    _Iyear_2004 _Iyear_2005 _Iyear_2006 _Iyear_2007)
  GMM-type (missing=0, separate instruments for each period unless collapsed)
    L(2/9).(L.kc openc cgnp)
------------------------------------------------------------------------------
Arellano-Bond test for AR(1) in first differences: z =  -2.94  Pr > z =  0.003
Arellano-Bond test for AR(2) in first differences: z =   0.23  Pr > z =  0.815
------------------------------------------------------------------------------
Sargan test of overid. restrictions: chi2(266)  = 465.53  Prob > chi2 =  0.000
  (Not robust, but not weakened by many instruments.)
Hansen test of overid. restrictions: chi2(266)  =  87.81  Prob > chi2 =  1.000
  (Robust, but can be weakened by many instruments.)
```

Given the relatively large number of time periods available, I have specified that the GMM instruments only be constructed for lags 2–9 to keep the number of instruments manageable. I am treating `openc` as a GMM-style instrument. The autoregressive coefficient is 0.648, and the `cgnp` coefficient is positive and significant. Although not shown, the test for joint significance of the time effects has p-value 0.0270.

We could also fit this model with the 'system GMM' estimator, which will be able to utilize one more observation per country in the level equation, and estimate a constant term in the relationship. I am treating lagged `openc` as a IV-style instrument in this specification.

```
. xtabond2 kc L.kc cgnp _I*, gmm(L.kc cgnp, lag(2 8)) iv(_I* L.openc) ///
> twostep robust nodiffsargan
Dynamic panel-data estimation, two-step system GMM
------------------------------------------------------------------------------
Group variable: iso                             Number of obs      =      1584
Time variable : year                            Number of groups   =        99
Number of instruments = 207                     Obs per group: min =        16
Wald chi2(17) =   8193.54                                      avg =     16.00
Prob > chi2   =     0.000                                      max =        16
------------------------------------------------------------------------------
             |              Corrected
          kc |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          kc |
         L1. |   .9452696   .0191167    49.45   0.000     .9078014    .9827377
             |
        cgnp |    .097109   .0436338     2.23   0.026     .0115882    .1826297
...
       _cons |  -6.091674    3.45096    -1.77   0.078    -12.85543     .672083
------------------------------------------------------------------------------
```

(continued)

```
----------------+----------------------------------------------------------------
Instruments for first differences equation
  Standard
    D.(_Iyear_1992 _Iyear_1993 _Iyear_1994 _Iyear_1995 _Iyear_1996 _Iyear_1997
    _Iyear_1998 _Iyear_1999 _Iyear_2000 _Iyear_2001 _Iyear_2002 _Iyear_2003
    _Iyear_2004 _Iyear_2005 _Iyear_2006 _Iyear_2007 L.openc)
  GMM-type (missing=0, separate instruments for each period unless collapsed)
    L(2/8).(L.kc cgnp)
Instruments for levels equation
  Standard
    _cons
    _Iyear_1992 _Iyear_1993 _Iyear_1994 _Iyear_1995 _Iyear_1996 _Iyear_1997
    _Iyear_1998 _Iyear_1999 _Iyear_2000 _Iyear_2001 _Iyear_2002 _Iyear_2003
    _Iyear_2004 _Iyear_2005 _Iyear_2006 _Iyear_2007 L.openc
  GMM-type (missing=0, separate instruments for each period unless collapsed)
    DL.(L.kc cgnp)
---------------------------------------------------------------------------------
Arellano-Bond test for AR(1) in first differences: z =  -3.29  Pr > z =  0.001
Arellano-Bond test for AR(2) in first differences: z =   0.42  Pr > z =  0.677
---------------------------------------------------------------------------------
Sargan test of overid. restrictions: chi2(189)   = 353.99  Prob > chi2 =  0.000
  (Not robust, but not weakened by many instruments.)
Hansen test of overid. restrictions: chi2(189)   =  88.59  Prob > chi2 =  1.000
  (Robust, but can be weakened by many instruments.)
```

Note that the autoregressive coefficient is much larger: 0.945 in this context. The `cgnp` coefficient is again positive and significant, but has a much smaller magnitude when the system GMM estimator is used.

We can also estimate the model using the forward orthogonal deviations (FOD) transformation of Arellano and Bover, as described in Roodman's paper. The first-difference transformation applied in DPD estimators has the unfortunate feature of magnifying any gaps in the data, as one period of missing data is replaced with two missing differences. FOD transforms each observation by subtracting the average of all *future* observations, which will be defined (regardless of gaps) for all but the last observation in each panel. To illustrate:

```
. xtabond2 kc L.kc cgnp _I*, gmm(L.kc cgnp, lag(2 8)) iv(_I* L.openc) ///
> twostep robust nodiffsargan orthog
Dynamic panel-data estimation, two-step system GMM
------------------------------------------------------------------------------
Group variable: iso                             Number of obs      =      1584
Time variable : year                            Number of groups   =        99
Number of instruments = 207                     Obs per group: min =        16
Wald chi2(17) =   8904.24                                      avg =     16.00
Prob > chi2   =     0.000                                      max =        16
------------------------------------------------------------------------------
             |              Corrected
          kc |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          kc |
         L1. |   .9550247   .0142928    66.82   0.000     .9270114     .983038
             |
        cgnp |   .0723786   .0339312     2.13   0.033     .0058746    .1388825
...
       _cons |  -4.329945   2.947738    -1.47   0.142    -10.10741    1.447515
------------------------------------------------------------------------------
```

(continued)

```
------------------------------------------------------------------------------
Instruments for orthogonal deviations equation
  Standard
    FOD.(_Iyear_1992 _Iyear_1993 _Iyear_1994 _Iyear_1995 _Iyear_1996
    _Iyear_1997 _Iyear_1998 _Iyear_1999 _Iyear_2000 _Iyear_2001 _Iyear_2002
    _Iyear_2003 _Iyear_2004 _Iyear_2005 _Iyear_2006 _Iyear_2007 L.openc)
  GMM-type (missing=0, separate instruments for each period unless collapsed)
    L(2/8).(L.kc cgnp)
Instruments for levels equation
  Standard
    _cons
    _Iyear_1992 _Iyear_1993 _Iyear_1994 _Iyear_1995 _Iyear_1996 _Iyear_1997
    _Iyear_1998 _Iyear_1999 _Iyear_2000 _Iyear_2001 _Iyear_2002 _Iyear_2003
    _Iyear_2004 _Iyear_2005 _Iyear_2006 _Iyear_2007 L.openc
  GMM-type (missing=0, separate instruments for each period unless collapsed)
    DL.(L.kc cgnp)
------------------------------------------------------------------------------
Arellano-Bond test for AR(1) in first differences: z =  -3.31  Pr > z =  0.001
Arellano-Bond test for AR(2) in first differences: z =   0.42  Pr > z =  0.674
------------------------------------------------------------------------------
Sargan test of overid. restrictions: chi2(189)   = 384.95  Prob > chi2 =  0.000
  (Not robust, but not weakened by many instruments.)
Hansen test of overid. restrictions: chi2(189)   =  83.69  Prob > chi2 =  1.000
  (Robust, but can be weakened by many instruments.)
```

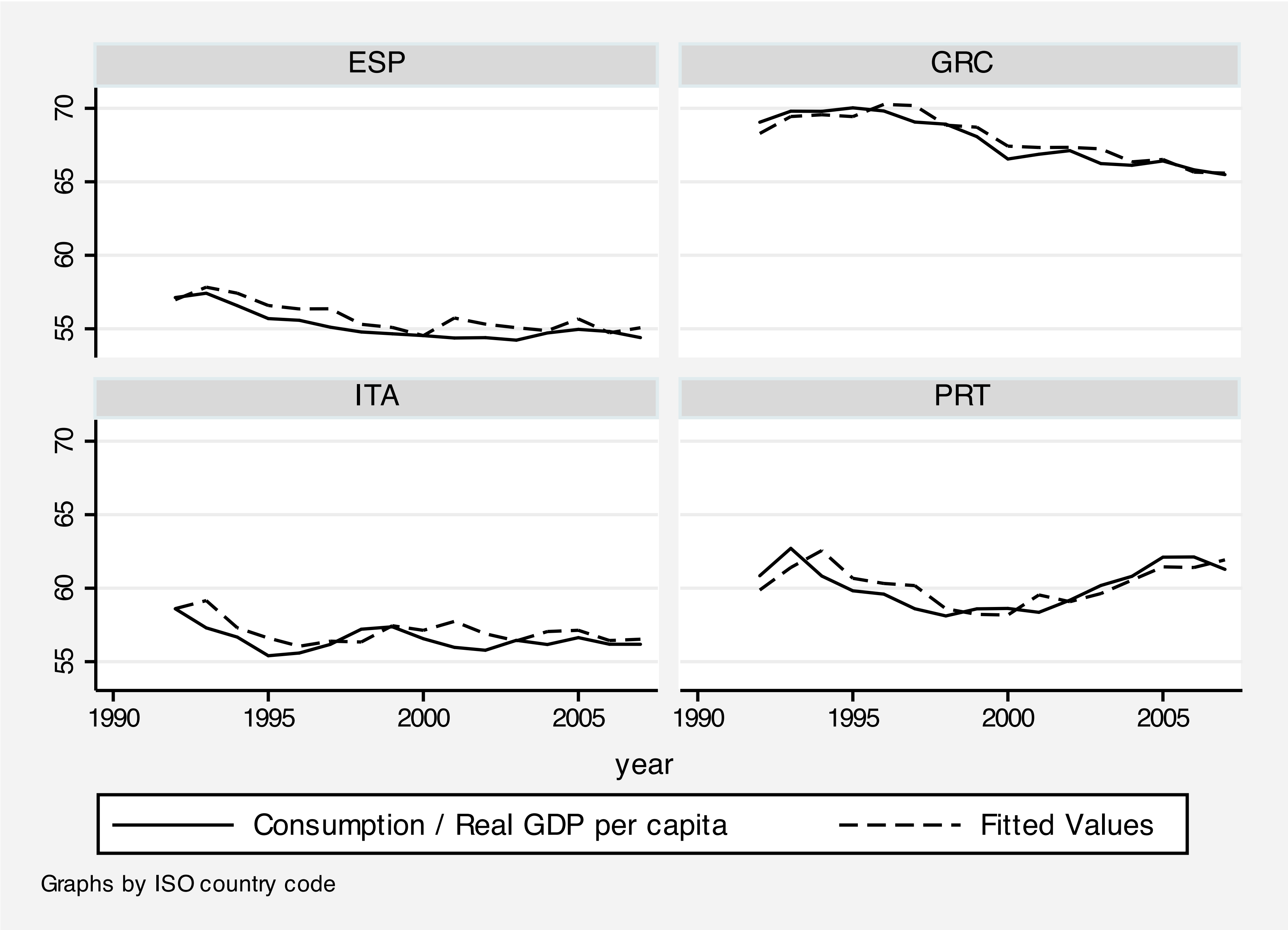
Using the FOD transformation, the autoregressive coefficient is a bit larger, and the `cgnp` coefficient a bit smaller, although its significance is retained.

After any DPD estimation command, we may save predicted values or residuals and graph them against the actual values:

```
. predict double kchat if inlist(country, "Italy", "Spain", "Greece", "Portugal
> ")
(option xb assumed; fitted values)
(1619 missing values generated)

. label var kc "Consumption / Real GDP per capita"

. xtline kc kchat if !mi(kchat), scheme(s2mono)
```


ESP
GRC
ITA
PRT
70
65
60
55
1990
1995
2000
2005
year
Consumption / Real GDP per capita
Fitted Values
Graphs by ISO country code

Although the DPD estimators are linear estimators, they are highly sensitive to the particular specification of the model and its instruments: more so in my experience than any other regression-based estimation approach.

There is no substitute for experimentation with the various parameters of the specification to ensure that your results are reasonably robust to variations in the instrument set and lags used. A very useful reference for DPD modeling is David Roodman's paper "How to do `xtabond2`" paper, freely downloadable from the *Stata Journal* via IDEAS or EconPapers.