CCHE CCES
Canadian Centre for Health Economics
Centre canadien en économie de la santé

Working Paper Series
Document de travail de la série

SOME IMPLICATIONS OF DYNAMIC MIS-SPECIFICATION FOR THE ARELLANO-BOND ESTIMATOR

Audrey Laporte, Adrian Rohit Dass, Brian S. Ferguson

Working Paper No: 160008

www.canadiancentreforhealtheconomics.ca

July 7, 2016

Canadian Centre for Health Economics
Centre canadien en économie de la santé
155 College Street
Toronto, Ontario

CCHE/CCES Working Paper No. 160008
July 7, 2016

Some Implications of Dynamic Mis-specification for the Arellano-Bond Estimator

Audrey Laporte[a,b,1], Adrian Rohit Dass[a,b], Brian S. Ferguson[b,c]

[a] *Institute of Health Policy, Management and Evaluation, University of Toronto*
[b] *Canadian Centre for Health Economics, Toronto, Canada*
[c] *Department of Economics, University of Guelph*

JEL Classification: C23; I12

Keywords: Arellano-Bond; dynamic panel

[1] Corresponding author. E-mail *audrey.laporte@utoronto.ca*.

I Introduction

While it is not the case that all Dynamic Panel Data (DPD) estimation assumes a single lag of the dependent variable, it is probably true that First Order lag structures dominate Second Order structures in the literature. It is also probably not unfair to suggest that the reasons for the presence of the lag tends not to be the subject of much theoretical discussion in the DPD literature. It can be argued that there are two factors at play here. On the practical side, since Arellano-Bond (AB) estimation (by which term we shall refer to a whole class of related GMM estimators) is designed for cases in which there is large N and small T, there is an inclination to economize on lagged dependent variables, especially if even longer lags are used as instruments for the included ones after differencing in an AB exercise. On the theoretical side, lagged dependent variables are probably most commonly interpreted as reflecting the dynamics of partial adjustment towards a long run (conditional on the values of the exogenous variables) equilibrium point. In that case, the inclusion of the lagged dependent variable is essentially intended to separate the long- from the short-run relation, and to prevent estimates of the long-run coefficients from being biased by the presence of un-controlled for short-run dynamics, especially in the case where there is considerable persistence in the short-run dynamics. In practical terms as well, there are not that many theoretical models that explicitly call for the presence of lags in excess of the first. Thus, there is a natural tendency to treat the presence of a single lag of the dependent variable as a maintained hypothesis, and not to test higher order lags.

There are, however, reasons to avoid complacency in the matter of the dynamic structure of DPD models, especially in applied health economics. One is the simple fact that longitudinal data sets are getting longer, so we have more scope to apply higher order dynamics. A second is that the assumption that dynamics represent partial adjustment towards an equilibrium point, while suitable for macro and market-level (i.e. commodity) models, is much more questionable in the case of individual-level models such as those that are likely to be estimated in health economics. In the context of aggregate level dynamic modeling, an equilibrium is not just a point towards which the system tends to converge (assuming it is dynamically stable), it also has the property that, once the system has reached it, it will sit there forever, so long as none of the exogenous variables change, causing the location of the equilibrium to shift. We refer to the system reaching its equilibrium, but when we are working with partial adjustment models as usually specified, using first order lags, it actually takes an infinite amount of time for the system to reach its equilibrium, although it can come arbitrarily close to equilibrium after a finite amount of time has elapsed. By contrast, individual level optimization problems, especially ones based on individual agents' inter-temporal optimization problems, will in general be finite horizon problems[2].

This is not to say that the finite-lived individual will never tend to converge towards a point, but that point will seldom be the equilibrium of a system. It is much more likely that the individual will converge towards a value of zero on some variable. An individual with no bequest motive, for example, is likely to aim to run his financial assets down to zero at the end of his expected horizon (probably with some allowance for the possibility that he will live longer than his mathematical life expectancy). A terminal asset value of zero, however, will in general not emerge as the long run target point of a partial adjustment model.

[2] An infinite horizon equilibrium, while it may technically be present in the math of the finite-lived individual's problem, makes no real sense as a target for the individual to be converging towards.

In fact, in individual-level inter-temporal optimization models, dynamics are much more likely to be intrinsic to the problem than they are to represent partial adjustment dynamics. Partial adjustment is generally taken to reflect the presence of costs to adjustment itself, which prevent the decision maker from jumping directly to the long run target value. Reduce the adjustment costs and the rate of convergence should increase. Eliminate the adjustment costs and the individual should always be at his optimum, even if the location of the optimum changes because the value of the exogenous variables change - in that case we move into a pure comparative statics world and the actual value of the choice variable will jump instantaneously to its new equilibrium value. In the case of the individual, however, the dynamics really summarize the shape of the optimal trajectory - in the case of the zero-endpoint target for assets, for example, there will be an optimal lifetime trajectory of saving and dissaving, and the instantaneous rate of each may change as time passes. This can result in the individual following a curved time-trajectory towards the ultimate endpoint, and while the saver might not think in terms of second order difference equations when she is making her lifetime asset accumulation plans, the Euler equations for her problem often translate into a second order difference equation in her choice variable.

We suggest in this paper that a great many models in health economics, including, but by no means limited to, models based on Grossman's concept of investment in health capital are likely to involve nonlinear trajectories. The possibility of curvature in the time-trajectory would by itself be sufficient reason for the applied health economist to work with a second order, rather than a first (i.e. linear) order structure.

Until recently, this has probably been a moot point. In the case of Grossman-type models, for example, the individual's level of health investment changes over time simply because of the passage of time. Lack of long, individual-level panel data sets has typically forced analysis to be done at the aggregate level, where we are dealing with transitional, or adjustment dynamics rather than intrinsic dynamics, because lagged adjustment processes are generally quite well represented by first order dynamics. As the availability of larger T panels has increased, however, health economists have been able to estimate individual-level theoretical models on individual-level data. We suggest that assuming first order dynamics in these cases raises the possibility of dynamic misspecification in DPD estimation. In this paper we conduct Monte Carlo experiments focused on the implications for AB estimation of mis-specification of the form of the intrinsic dynamics likely to be present in individual models of health and health behaviour. In Section II we discuss the difference between transition dynamics and intrinsic dynamics and the implications of this distinction for econometric modeling. In the Section III we outline a theoretical model of individual behavior that commonly underlies estimation of individual investment in health decisions, drawing out the detail of where the intrinsic dynamics manifests within it. Section IV outlines the AB approach in the context of mis-specification of intrinsic dynamics, with a particular focus on the case where a First Order Difference equation (FODE) is estimated but the true intrinsic dynamics are in fact generated by a Second Order Difference equation (SODE). Section V describes the Monte Carlo (MC) experiments employed to investigate the consequences of the mis-specification for bias in the X coefficients, the lagged variables and the performance of test statistics in identifying the mis-specification. Section VI reports the results of the MC experiments. Section VII concludes and highlights areas for future research.

II Dynamics - a theoretical perspective

Transitional dynamics in micro and macro modelling:

It is important to consider the difference between dynamics at the aggregate level and dynamics at the individual level, and within each level between finite and infinite horizon models. Macrodynamics typically deals with the process of transition from one equilibrium to another or, more to the point, the adjustment of the macro system to its equilibrium after a shock. That shock could be a deep structural shock, meaning one which shifts the location of the equilibrium, in which case it is the process by which the actual values of the system move from the values which were associated with the old equilibrium (on the assumption that the system was at the old equilibrium when the shock occurred) to those associated with the new equilibrium. Recall the simple undergraduate textbook-level illustration of the IS-LM model when the level of government spending changes. The immediate effect is to shift the IS curve, creating a new IS-LM equilibrium point, but since the IS curve is an equilibrium curve, not a behavioural relation, we must also model the way the actual values of income (Y) and the interest rate (r) converge on their new equilibrium values. So long as the model is dynamically stable, the actual values of Y and r will converge on their new equilibrium values, following a trajectory that is determined by the assumptions made about the speeds of adjustment in the respective markets. If, for example, the money market clears so quickly that it can be regarded as always being in equilibrium, the transition path will lie along the LM curve. If both markets were to clear extremely quickly, we would virtually be in a comparative statics world.

Similarly, still at the macro level, we can have shocks that kick the system away from its equilibrium point without moving the location of the equilibrium itself. This could be the case of a classical economics downturn, after which the economy will return to its full employment equilibrium. As in the first example, the dynamics observed are *transitional* dynamics, and can be thought of as lagged adjustment from the initial point of the system (meaning the point at which the system is sitting immediately after the shock, when the problem begins) to the equilibrium.

Transitional dynamics are not absent in microeconomics. Consider the profit maximizing firm's problem of determining its optimal level of capital input. The first order conditions for the problem can be used to define the individual firm's optimal capital input, but absent adjustment costs, the firm will prefer to jump to the optimal capital level. Looked at as a time trajectory, this would translate into an infinite rate of investment over a vanishingly small interval. In order to get a more realistic investment time path we must assume adjustment and transition costs of some sort or other. Sometimes these costs can be explicitly modeled, on theoretical principles, and other times it is convenient simply to assume a lagged adjustment process and to put lagged values of capital in the equation explaining the current value of capital.

Similarly, while the profit maximizing level of employment will be at the level at which the value of the marginal product of labour equals the wage of labour, there may be adjustment costs and lags which prevent the firm from jumping instantaneously to the new, optimal level of labour in response to a change in the wage rate. This can be thought of as the firm-level counterpart of the macro concept of frictional unemployment. The length of the lag will depend on the labour market institutions and conditions within which the firm is operating: if the labour market can be made more efficient, the lag time will be less. If labour market institutions are so efficient as to permit the firm to jump immediately to the new profit maximizing level of employment, the lag terms will vanish altogether. Thus, when Arellano and Bond (1991), in their development of the m1 and m2 statistics[3], specify the employment function for a firm as involving two lags of the dependent variable (the level of employment), they are allowing for adjustment costs.

III Intrinsic dynamics in micro modeling

When we consider micro-level dynamic problems, we need to give careful consideration to the time horizon of the problem. Macro-dynamic problems tend intrinsically to be infinite horizon problems, since it makes more sense to assume that any given macro-economy will be around forever than it does to assume that it will come to an end at some predetermined date. The inter-temporal decisions made by individuals, on the other hand, are much more likely to be finite horizon problems[4]. When we consider the individual's optimal lifetime trajectory of accumulation of financial assets, for example, we typically find that most individuals follow a path which involves a period of dissaving, during the family formation years, followed by a period of asset accumulation during the prime labour force years, and a further period of asset decumulation during retirement. In some cases the endpoint of the individual's lifetime saving problem may be determined by a bequest target, and, if we choose a specific form for the bequest function we can in fact make the individual saver behave as if she intended to live forever - this is how representative agent macro models combine the infinite horizon which makes sense for macro problems with the assumption that macro behavioural relations such as the consumption function are derived from individual optimizing behavior. More commonly, though, the logical horizon is finite and the individual will follow an inverted-U lifetime saving time path with its endpoint determined by the target bequest level. The inverted-U time path should be thought of as an example of dynamics which are *intrinsic* to the solution of the optimization problem, rather than as lagged adjustment process, since it will be determined by the first order conditions for the problem even in the absence of adjustment costs and factors which could cause the adjustment to be lagged. When this saving path is represented by a difference equation in saving, the likelihood that it will need to change direction over time dictates that it should not be represented by a First Order Difference Equation (FODE), but rather by a Second Order Difference Equation (SODE).

Individual level investment problems, whether investment in financial assets or investment in health, involve the decision maker using a control variable, some form of investment, to guide the accumulation over time of a state, or stock variable, such as the stock of health capital or accumulated financial wealth. In each case the stock variable itself can be linked to the individual's future utility - health capital directly and financial capital through the future consumption it permits. The individual's actual dynamic decision rules, derived from her inter-temporal optimization problem, should ideally be represented econometrically by a dynamic system of interrelated FODEs in which the current values of the each of the state and control variables depend on their own past values and the current values of the other variable.

[3]These are the common tests for serial correlation in DPD estimation.

[4]Note that the dynamics of the market will be different from those of the individual, even though market-level behavior is a result of the decisions of those individuals. In part this is due to the aggregation conditions which apply at any point in time. In addition, the market can be thought of as infinitely-lived, even though individuals will only be in it for a finite period, because departing individuals will be replaced by new arrivals.

In general terms, letting S stand for the state, or stock variable and C for the control variable, we can write

$$S_t = g(C_{t-1}, S_{t-1}) \tag{1}$$

where the function $g(C_{t-1}, S_{t-1})$ summarizes all of the factors which affect the rate at which S cumulates: in the case of a physical investment process, and the process of investment in health capital, these would include the production relation by which C adds to S and the rate at which S depreciates in the absence of positive S. The g(.) function, then, can be thought of as a purely technical expression. The optimal behavioural rule for the problem can be written as:

$$C_t = h(C_{t-1}, S_{t-1}, X_t, S_T) \tag{2}$$

Where the decision with regards to the current level of C depends on the past level of C (through some form of Euler equations, taking account of the opportunity cost of C), the previous period's actual level of S, and S_T, which can be taken to be the optimal terminal value of S, to be attained at some finite, end of horizon, time T and which may have been chosen as part of the optimization problem itself. In equation (2), X_t stands for the other, exogenous variables that might play into the individual's decision - prices and income, for example. Thus the current period's value of C depends on where the individual is, in terms of their value of S, where they want to end up at the finite end-time of the problem, the rules which explain the growth of S and the individual's preferences, which determine the marginal opportunity cost of C. This way of expressing the solution to the optimization problem yields a pair of interrelated FODEs in C and S, and suggests that, when trying to analyze econometrically the individual's choices with regards to C, it is necessary to take account of the relation between C and S. Ideally it would also be useful to take account of S_T, since the two stock elements in the equation for C relate to where the individual is, where she wants to end up and how long she has to reach that point, but in most cases that information is not available. The individual is presumed to know where she is heading, and to follow an optimizing trajectory to that point, but from the perspective of the econometrician, S_T is an unobservable element of individual heterogeneity, of the sort that causes problems for OLS estimation of the dynamics of C.

Estimating this pair of equations would require a simultaneous panel data estimation process. Further, it may well be, especially in health economics applications, that the data on S (i.e. on the individual's stock of health capital) may not be available. This would seem likely to cause significant omitted variable bias problems for empirical investigation of individual health investment behavior.

In the theoretical literature on dynamic optimization, however, it is argued that the optimal trajectory for the individual's capital accumulation and decumulation problem (since a finite horizon problem will involve both) can be represented by a higher order difference equation in the control variable alone (See, for example, Ferguson and Lim (2003)). In essence, having discovered a policy rule which relates current investment to current assets, which are in turn the result of past investment and which are going to be decumulated through a dis-investment process in order to permit future consumption to be higher than would otherwise be the case, the policy rule can be substituted into the investment problem to obtain a characterization of the optimal time path of investment in which the current stock of the capital asset relevant to the problem has been optimally substituted for by past values of the investment variable. In terms of our notation, the theoretical literature argues that substitution between equations (1) and (2) allows for estimation of an equation of the form

$$C_t = h(C_{t-1}, C_{t-2}, X) \tag{3}$$

While this substitution involves the loss of a certain amount of information, in that it may not be possible to identify all of the structural parameters of (1) and (2) when estimating (3), it is still possible to explain the individual's time trajectory of C, and because of the presence of the X variables in (3), it is possible to show how the optimal time-trajectory shifts when the values of the exogenous variables change. While in general, the parameters of the production relation (1) cannot be identified from (3), as long as the technology through which S changes in response to changes in C remains unaltered over time, the crucial parameters of the individual's choice rule can be estimated, and this is the behavioural equation of primary interest to health economists.

This type of substitution of past values of the control variable for current values of the stock variable was once common in macroeconomic modeling of the aggregate physical capital stock, where the current stock of aggregate capital, being unobservable, was proxied by a distributed lag over investment. In the case of individual level problems, a similar form of substitution can be undertaken, with the precise form of the lag process emerging from the individual's finite-horizon inter-temporal optimization problem. Thus in a problem which involves investment and the accumulation of some form of capital, which includes health capital in Grossman type problems and addiction capital in Becker-Murphy type problems, we can choose between estimating a pair of interrelated FODEs, one in the investment variable and one in the capital variable, or estimating a higher order equation in just one of the variables alone. In panel data applications it is common to estimate a first order system in just one of the variables (although Arellano (2003), for example, sets out the GMM equations for the Rational Addiction model) by analogy with infinite horizon lagged adjustment problems, but the nature of the underlying problem often calls for a different dynamic structure.

Analyzing the intrinsic dynamics within the individual-level economic model

To see how intrinsic dynamics translate into the time-trajectory of the dependent variable in an individual-level economic model, consider the case of a simple FODE:

$$Y_t = \alpha_0 + \alpha_1 Y_{t-1} + \beta X_t \tag{4}$$

The solution function for this equation, as a function of time, is

$$Y_t = A\lambda^t + Y^* \tag{5}$$

Where Y^* is the equilibrium of the system. Y^* is part of the solution function whether or not the system is going to converge on it - for example, if the equilibrium is unstable, so that the actual value of Y diverges from its equilibrium value, the equilibrium value is still part of the formal expression of the solution, showing how rapidly Y diverges from its equilibrium as time passes. The actual value of Y^* will depend on the value of X, the exogenous variable(s) of the problem. The term λ is the single root of the FODE, and determines whether the equilibrium is stable or not. The root is a real constant, and in most economic examples of FODEs is positive[5]. Since a positive root is the most common case in economic analysis, we shall in general maintain that assumption here. The equilibrium point is dynamically stable if λ is less than 1 in absolute value. The term A is a constant which depends on the conditions of the problem and is solved for as part of the solution process - it can be positive, negative or zero, but its value only changes if the basic conditions of the problem change. In effect, A anchors the position of the optimal time trajectory.

Looking at the slope of the trajectory with respect to elapsed time, we have

$$\partial Y_t/\partial t = [ln\lambda]A\lambda^t \tag{6}$$

If we assume that λ is a positive fraction, so the system is dynamically stable (we set aside the case of a unit root in this discussion) then lnλ will be negative and the sign of the slope of the time trajectory will depend on the sign of A. If A is positive, Y will be diminishing over time and if A is negative, Y will be increasing over time. Whichever the case, it is clear that, since A and λ do not change over time (and since we have assumed that λ is positive), the time trajectory of Y is monotonic - it will not change direction.

When working with a SODE of the form:

$$Y_t = \alpha_0 + \alpha_1 Y_{t-1} + \alpha_2 Y_{t-2} + \beta X_t \tag{7}$$

The solution function for Y as a function of t is

$$Y_t = A_1\lambda_1^t + A_2\lambda_2^t + Y^* \tag{8}$$

Where we now have two roots, λ_1 and λ_2, and two undetermined constants, A_1 and A_2. This structure clearly allows a wider range of time paths of Y, including cyclical ones, when the roots constitute a complex conjugate pair. We are again concerned with the case where the λs are real, in general positive, and less than 1 in absolute value. Looking at the slope of the trajectory of Y with respect to time yields:

[5]One well-known exception to this generalization is the cobweb model of dynamic pricing, in which case the root will be negative.

$$\partial Y_t / \partial t = [ln\lambda_1] A_1 \lambda_1^t + [ln\lambda_2] A_2 \lambda_2^t \quad (9)$$

Clearly it will now be possible for the trajectory to change direction, even in the case where both roots are positive fractions and the system is therefore dynamically stable, depending on the signs of the A terms. If the As are opposite in sign and there is a value of t such that

$$[ln\lambda_1] A1 \lambda_1^t + [ln\lambda_2] A_2 \lambda_2^t = 0 \quad (10)$$

then the slope of the time trajectory can reverse. This is not the same as the case of complex roots, which will induce repeating cycles. When the roots are both real, positive and less than 1 in absolute value the trajectory can change direction only once, but that is all that is needed for an inverted-U trajectory with respect to time.

Since health economics frequently deals with inter-temporal optimization problems at the individual level, it seems prudent that applied health economists at least consider the likelihood that that they are working in a world of second order dynamics. By extension, it would seem to be important to also consider the econometric implications of working with a first order DPD structure when the true dynamics are more appropriately characterized by second order dynamics. This latter question is the focus of the present paper. In particular, we are concerned with the implications of mis-specified dynamics and with whether there is any way of identifying whether the dynamics in a particular DPD exercise have been mis-specified.

IV AB estimation and dynamic mis-specification

That mis-specified dynamics could potentially be a serious issue in AB-type estimation follows from the fact that the GMM approach typically uses lags of the dependent variable as instruments. The usual estimation structure starts from a FODE of the form:

$$Y_{it} = \alpha_0 + \alpha_1 Y_{it-1} + \beta Z_{it} + \eta_i + \epsilon_t \quad (11)$$

where i indexes the individual and t indexes time. The term α_0 is a common intercept while η_i is an individual-specific intercept. In small T-large N panels, the presence of η_i in the data generating process means that OLS estimates of α_1 are likely to be biased: individuals with large positive η_i terms will tend to consume larger quantities of Y in all periods than will individuals with small individual intercepts, creating a persistence in Y over time which will tend to manifest in OLS estimation as a significant positive value of α_1, even when there is no habit formation element involved.

Since the inclusion of individual dummy variables does not resolve the problem, the usual approach is to difference (11) to eliminate the time-invariant terms for each individual:

$$\Delta Y_{it} = \alpha_1 \Delta Y_{it-1} + \beta \Delta Z_{it} + \Delta\epsilon_t \tag{12}$$

In (12), $\Delta Y_{it-1} = Y_{it-1} - Y_{it-2}$, and $\Delta\epsilon_t = \epsilon_t - \epsilon_{t-1}$, and since Y_{it-1} depends on ϵ_{t-1}, the differencing operation has introduced an endogeneity problem into the estimation. This is typically dealt with by instrumenting ΔY_{it-1}: the fact that, by construction, Y_{it-2} is correlated with ΔY_{it-1} but not with $\Delta\epsilon_t$ explains the use of lagged values of Y in the instrumenting equation.

Now consider the case where the true DGP is a SODE:

$$Y_{it} = \alpha_0 + \alpha_1 Y_{it-1} + \alpha_2 Y_{it-2} + \beta Z_{it} + \eta_i + \epsilon_t \tag{13}$$

If it is assumed that the true DGP is a FODE and the estimating equation is mis-specified, taking the DGP to be (11) rather than (13), there is an omitted variable problem, and the effect of Y_{it-2}, the omitted variable, will be included, along with the true disturbance term ϵ_t, in the regression residual term, giving:

$$Y_{it} = \alpha_0 + \alpha_1 Y_{it-1} + \beta Z_{it} + \eta_i + [\epsilon_t + \alpha_2 Y_{it-2}] \tag{14}$$

Differencing (14) yields:

$$\Delta Y_{it} = \alpha_1 \Delta Y_{it-1} + \beta \Delta Z_{it} + [\Delta\epsilon_t + \alpha_2 \Delta Y_{it-2}] \tag{15}$$

As before there is a problem of correlation between $\Delta\epsilon_t$ and ΔY_{it-1}, but now there is also a correlation between ΔY_{it-2} and ΔY_{it-1}, and further, if Y_{it-2} is used in the instrumenting step, there will be correlation between Y_{it-2} and the ΔY_{it-2} element in the residual. In fact, since $\Delta Y_{it-2} = Y_{it-2} - Y_{it-3}$, Y_{it-3} should not be used as an instrument, but the instrument set should start at Y_{t-4} which, depending on the DGP and the true persistence in T over time, could be a very weak instrument. Clearly the potential exists for some fairly significant consequences to arise from mis-specified dynamics in DPD estimation.

Problems associated with mis-specified dynamics are not uncommon in other areas of economics. The macro-econometric literature on unit roots and co-integration developed in part out of a realization that autocorrelation in the disturbance term was likely to reflect the omission of a lagged dependent variable as it was to reflect true structural autocorrelation in the disturbance term. This precedent raises the question of whether tests for serial correlation in the residuals of an AB estimated equation can be read as signals of the presence of mis-specified dynamics. In particular, we are interested in whether the m1 and m2 statistics (Arellano and Bond (1991)), which are the common tests for serial correlation in DPD estimation, can be read as signals of the presence of mis-specified dynamics. The issue here is slightly more involved than it was in the case of the macro-dynamic literature: the differencing process associated with going from equation (1) to equation

(2) above involves differencing the disturbance term, and the m-statistics are generated using the residuals, $\Delta\epsilon_t$, of equation (2), and since $\Delta\epsilon_t = \epsilon_t - \epsilon_{t-1}$, and $\Delta\epsilon_{t-1} = \epsilon_{t-1} - \epsilon_{t-2}$, the residuals will display moving average serial correlation. Indeed, a signal by the m1 statistic that there is serial correlation in the residuals is taken as evidence that the residuals of the equation in levels are non-correlated. Thus we must expect to have to look to the m2 statistic, for second order serial correlation in the residuals, for our test of whether there is serial correlation in the residuals of the estimated equation in levels resulting from mis-specified dynamics.

That mis-specification can cause problems for estimation is not a novel result, of course, but the literature does not give a lot of attention to issues of dynamic mis-specification in DPD estimation. The most complete discussion of dynamic mis-specification is probably that by Lee (2012), but that discussion deals with the implications of mis-specified dynamics for the Nickell (1981) bias problem, and not directly with the issue dealt with here. Jung (2005) mentions mis-specified dynamics as a possible source of autocorrelation in the residuals of a DPD estimation problem, but does not discuss the issue beyond that, preferring to focus on cases in which the disturbance term has a genuine AR(1) structure. Other discussions of the nature of the dynamics in DPD problems tend to focus on the issue of a unit root, or near unit root, in the FODE for what we have labelled the S variable - Kruiniger (2009) and Phillips (2014), for example, and this strand of the literature is related to the panel co-integration literature (see, for example, Choi (2015)). Within the Panel co-integration literature there is some work dealing with testing for co-integration in panel ARDL models (see, for example, Pesaran (2015)) and the use of error-correction forms of ARDLs (Westerlund (2007)), but these models typically work in the lagged adjustment framework and use first order ARDLs, and so do not deal with quite the same issue being dealt with here.

V Monte Carlo Simulation

The Monte Carlo experiments that are conducted are focused on the implications of mis-specified dynamics in AB estimation. In each of the experiments reported here the system is dynamically stable. In the experiments reported below, the true DGP is a Second Order Difference equation (SODE) structure but the estimated equation is a First Order Difference equation (FODE). The coefficients on the two lags of the dependent variable are varied across the experimental DGPs, but the coefficient on the exogenous variable is kept constant at 1. For each run the m1 and m2 statistics are generated, in order to get a sense of how well those test statistics serve as signals of the presence of mis-specified dynamics. Since the m1 and m2 statistics are generated using the residuals from the differenced equations, the expectation is that the m1 statistic will detect the MA structure which the differencing imparts to the disturbances, even when there is no autocorrelation in the disturbance terms in levels. In the experiments, the dependent variable yit was generated from the following equation:

$$y_{it} = \alpha_1 y_{i(t-1)} + \alpha_2 y_{i(t-2)} + \beta x_{it} + \eta_i + \epsilon_{it} \tag{16}$$

where i=1,...,N, t=1,...,T+10, $\eta_i \sim$ i.i.d. N(0, 1), $\epsilon_{it} \sim$ i.i.d. N(0, 1), and $y_{i1} = y_{i2} = 0$. Following Arrelano & Bond (1991), the first 10 cross-sections are discarded, leaving a panel of T=9 and N=100. Each experiment was run using 500 Monte Carlo replications and $\beta = 1$ in all experiments.

The exogenous variable xit was generated from the following equation:

$$x_{it} = \rho x_{i(t-1)} + \mu_{it} \tag{17}$$

where $\rho = 0.85$ and $\mu_{it} \sim$ i.i.d N(0, 0.81), independent of η_i and ϵ_{it}. The values of x_{it} were kept fixed across replications in each of the experiments.

The Monte Carlo routine was written in R version 3.2.2 (R Core Team, 2013). Pooled OLS estimation was performed using the 'lm' command in R. One-step GMM estimation using first differences was estimated using the 'plm' package (Croissant & Millo, (2008)).

VI Monte Carlo Results

This section reports the results of Monte Carlo experiments of AB-type estimation of SODE dynamics. In performing the GMM estimation it is assumed that the investigator does not know that the dynamics of the problem are mis-specified, and so uses as an instrument a lag of the dependent variable that is appropriate to the case where the true DGP is a FODE, not a SODE. The AB m1 and m2 test statistics are generated, since we are interested in whether, by analogy with the macro-dynamic literature on econometric mis-specification of the AR structure of dynamics, these widely used test statistics for the presence of autocorrelation in AB-type estimation exercises will provide a signal as to whether there is a problem of mis-specified dynamics. Each table below reports the AB-type GMM results for both correctly specified and mis-specified cases, (where mis-specified refers to having run a FODE estimating equation where the true DGP is a SODE) along with the Monte Carlo standard deviations for the coefficients. For each of the m_1 and m_2 statistics, the percentage of times the test rejects the null of no autocorrelation is also reported.

In assessing the results it is important to keep track of the number of problems we face. There are in fact three problems intermingled in the estimation results. First, we are assuming that the researcher has mis-specified the dynamics of the problem as FODE rather than the true SODE. This could happen, for example, if she assumed that the dynamics were of the lagged adjustment form rather than being intrinsic to the problem, and failed to allow for the possibility that the solution time trajectory might be curved. This is the initial omitted variable problem.

Second, we have Nickell's (1981) problem: we want to estimate the equation using individual level panel data, but each individual in the panel can have a unique intercept, and this can bias the estimate of the coefficient on the lags - a problem which would arise even if the dynamic structure had been specified properly.

Our third problem arises from the fact that the dynamic structure has not been specified properly. Because we are dealing with dynamic panel data, we difference to resolve the Nickell problem, and use AB-type GMM estimation to estimate the coefficients of the presumed equation. The residual term of the GMM estimation problem has two issues: one is the omitted variable factor, since the effect of Y_{t-2} is now included in the residuals, the other is an erroneous instrument problem since, having differenced the equation for estimation purposes, we select a lag of Y to use as an instrument based on what we think the dynamic structure is, not on what it actually is. Thus, we do not choose a high enough lag for the instrument and introduce another source of endogeneity bias. The first set of Monte Carlo results are set out in Table 1 below:

Table 1: True α_1 = 0.7, α_2 = -0.10, β = 1, λ_1 = 0.5, λ_2 = 0.2

	GMM		GMM Mis-specified	
	Beta	SD Beta	Beta	SD Beta
α_1	0.671	0.067	0.606	0.051
α_2	-0.099	0.035	N/A	N/A
β	0.998	0.065	1.031	0.059
First order serial correlation	1	N/A	1	N/A
Second order serial correlation	0.064	N/A	0.254	N/A

The bias in the estimation of the coefficient on Y_{t-1} is not large and the coefficient on the exogenous variable is well estimated. The m_1 test rejects the null of no serial correlation 100% of the time, as it should. The m_2 test rejects the null of no second order serial correlation only about 25% of the time, so it cannot be seen as sending a clear signal that there might be a problem. The second Monte Carlo run is set out in Table 2:

Table 2: True α_1 = 0.7, α_2 = -0.06, β = 1, λ_1 = 0.6, λ_2 = 0.1

	GMM		GMM Mis-specified	
	Beta	SD Beta	Beta	SD Beta
α_1	0.666	0.072	0.629	0.055
α_2	-0.057	0.036	N/A	N/A
β	0.998	0.064	1.02	0.059
First order serial correlation	1	N/A	1	N/A
Second order serial correlation	0.076	N/A	0.134	N/A

Here again the erroneous omission of the second lag cannot be said to be causing major problems for the estimation of the coefficients on the included variables. Table 3 below raises a more interesting issue:

Table 3: True α_1 = 1.5, α_2 = -0.56, β = 1, λ_1 = 0.8, λ_2 = 0.7

	GMM		GMM Mis-specified	
	Beta	SD Beta	Beta	SD Beta
α_1	1.46	0.069	0.796	0.028
α_2	-0.53	0.054	N/A	N/A
β	0.99	0.068	1.05	0.089
First order serial correlation	1	N/A	0.26	N/A
Second order serial correlation	0.05	N/A	0.14	N/A

In Table 3 the estimated value of the coefficient on Y_{t-1} is quite different from its true value. We might, therefore, argue that we have a much more serious bias problem in this case than in the previous two. On the other hand, it is not at all clear that we want the estimate of the value of the coefficient on Y_{t-1} to be close to its true value in the face of the omission of Y_{t-2} from the right hand side. The true DGP is a second order difference equation with two stable, positive roots. The mis-specified estimating equation is a first order difference equation. If the estimated coefficient on Y_{t-1} were an unbiased estimate of its true value in the DGP, it would be close to 1.5, and our estimated FODE would be unstable, even though the true DGP was a stable SODE. We see the same issue in Table 4 below:

Table 4: True α_1 = 1.1, α_2 = -0.28, β = 1, λ_1 = 0.7, λ_2 = 0.4

	GMM		GMM Mis-specified	
	Beta	SD Beta	Beta	SD Beta
α_1	1.06	0.074	0.725	0.048
α_2	-0.261	0.046	N/A	N/A
β	0.996	0.065	1.11	0.06
First order serial correlation	1	N/A	1	N/A
Second order serial correlation	0.05	N/A	0.66	N/A

Here, again if in estimating our misspecified equation we were able to obtain an unbiased estimate of the true DGP coefficient on Y_{t-1}, we would conclude that the true DGP was an unstable FODE whereas in fact it is a stable SODE.

In a FODE there is only a single root determining the dynamics of Y, and that root is the same as the coefficient on Y_{t-1}. In the cases shown in Tables 3 and 4 above, while the coefficient on Y_{t-1} in the misspecified equation is notably different from the coefficient on Y_{t-1} in the DGP, it is in fact quite close to the value of the larger root in the true SODE. In the case of a SODE with stable roots, the larger of the two roots is the dominant root, so it appears that, while the misspecified equation is yielding a biased estimate of the coefficient on Y_{t-1}, it is in effect providing a much less biased estimate of the true dynamics of Y. We see the same effect in Table 5 below:

Table 5: True α_1 = 1.7, α_2 = -0.72, β = 1, λ_1 = 0.9, λ_2 = 0.8

	GMM		GMM Mis-specified	
	Beta	SD Beta	Beta	SD Beta
α_1	1.659	0.072	0.865	0.016
α_2	-0.685	0.063	N/A	N/A
β	0.982	0.073	0.742	0.066
First order serial correlation	1	N/A	1	N/A
Second order serial correlation	0.046	N/A	0.85	N/A

Here, while the estimated coefficient on Y_{t-1} in the mis-specified equation is only about half of the true DGP coefficient on Y_{t-1}, it is extremely close to the value of the larger of the two roots.

To investigate this further, we ran a number of Monte Carlo experiments using as the true DGP a SODE with complex roots of modulus less than one - i.e. the true DGPs in the following cases display stable cyclical behavior. In addition to calculating the roots we have, for each case in which the roots are complex, calculated the periodicity of the cycle (see Baumol (1970) for the procedure).

Table 6: True α_1 = 1.4, α_2 = -0.6, β = 1, $\lambda_{1,2}$ = 0.7 ± 0.33i, modulus = 0.77, periodicity = 14

	GMM		GMM Mis-specified	
	Beta	SD Beta	Beta	SD Beta
α_1	1.385	0.039	0.752	0.032
α_2	-0.591	0.031	N/A	N/A
β	0.998	0.065	1.367	0.074
First order serial correlation	1	N/A	0.094	N/A
Second order serial correlation	0.04	N/A	0.67	N/A

Table 7: True α_1 = 0.6, α_2 = -0.12, β = 1, $\lambda_{1,2}$ = 0.3 ± 0.173i, modulus = 0.346, periodicity = 12

	GMM		GMM Mis-specified	
	Beta	SD Beta	Beta	SD Beta
α_1	0.576	0.062	0.527	0.048
α_2	-0.122	0.034	N/A	N/A
β	0.999	0.065	1.022	0.061
First order serial correlation	1	N/A	1	N/A
Second order serial correlation	0.066	N/A	0.414	N/A

Table 8: True α_1 = 0.77, α_2 = -0.40, β = 1, $\lambda_{1,2}$ = 0.385 ± 0.5i, modulus = 0.63, periodicity = 7

	GMM		GMM Mis-specified	
	Beta	SD Beta	Beta	SD Beta
α_1	0.758	0.045	0.539	0.035
α_2	-0.402	0.03	N/A	N/A
β	0.998	0.065	1.072	0.068
First order serial correlation	1	N/A	0.998	N/A
Second order serial correlation	0.06	N/A	1	N/A

These three tables show quite mixed results in terms of the value of the coefficient on Y_{t-1}, the coefficient on the exogenous variable, and the signals being sent by the m_1 and m_2 statistics. Overall, while the omitted variable bias issue seems to be least notable in the case of the coefficient on the exogenous variable, there are cases where the estimated value of β is quite different from the true.

We noted above that we face three issues in our estimation: the misspecification, Nickel's (1981) problem, and the fact that we are using an incorrect instrument in the AB estimation. In our next set of Monte Carlo runs, in order to try and get a sense of the importance of the various factors, we add a set of results to the tables. In these tables, we consider would happen if we were to try and eliminate the instrumenting problem by using Y_{t-4} as the instrument. Clearly we are making an odd assumption - there is no reason for the researcher to do this unless she believes that she has mis-specified the dynamics - but it gives us a bit of a picture of the intermingled sources of bias. In order to do this we have increased T from 8 to 9. In each of the following tables, then, the final two sets of columns give the results of AB estimation in the case where the researcher does not realize that she is using an invalid instrument and, in the last set of columns, the results of Monte Carlos where the equation is still mis-specified but, oddly, the researcher uses an appropriate, if rather weak, instrument. Table 9 below corresponds to Table 1 above:

Table 9: True α_1 = 0.7, α_2 = -0.1, β = 1, λ_1 = 0.5, λ_2 = 0.2

	GMM New		GMM New Mis-specified		GMM Mis-specified IV 4 lags+	
Coefficient	Beta	SD Beta	Beta	SD Beta	Beta	SD Beta
α_1	0.681	0.055	0.589	0.042	0.597	0.055
α_2	-0.099	0.033	N/A	N/A	N/A	N/A
β	1.003	0.053	1.05	0.047	1	0.058
First order serial correlation	1	N/A	1	N/A	1	N/A
Second order serial correlation	0.078	N/A	0.3	N/A	0.214	N/A

Here we see that there is a bit of sensitivity to the increase in T by one, but the results are not, overall much affected. Corresponding to Table 5 we have Table 10:

Table 10: True α_1 = 1.7, α_2 = -0.72, β = 1, λ_1 = 0.9, λ_2 = 0.8

	GMM New		GMM New Mis-specified		GMM Mis-specified IV 4 lags+	
Coefficient	Beta	SD Beta	Beta	SD Beta	Beta	SD Beta
α_1	1.674	0.046	0.87	0.013	0.787	0.024
α_2	-0.699	0.041	N/A	N/A	N/A	N/A
β	0.985	0.059	0.905	0.073	0.779	0.066
First order serial correlation	1	N/A	1	N/A	1	N/A
Second order serial correlation	0.056	N/A	0.99	N/A	1	N/A

Things become somewhat more interesting in our next set of Monte Carlo runs, when considering the cases with complex roots. These runs do not have counterparts in our earlier tables.

Table 11: True α_1 = 0.6, α_2 = -0.6, β = 1, $\lambda_{1,2}$ = 0.3 ± 0.71i, modulus = 0.77, periodicity = 4.83

	GMM New		GMM New Mis-specified		GMM Mis-specified IV 4 lags+	
	Beta	SD Beta	Beta	SD Beta	Beta	SD Beta
α_1	0.598	0.029	0.411	0.024	-0.139	0.113
α_2	-0.602	0.026	NA	NA	NA	NA
β	0.996	0.049	0.907	0.066	0.85	0.082
First order serial correlation	1	NA	0.122	NA	0.62	NA
Second order serial correlation	0.07	NA	1	NA	1	NA

Table 12: True α_1 = 0.5, α_2 = -0.5, β = 1, $\lambda_{1,2}$ = 0.25 ± 0.66i, modulus = 0.71, periodicity = 4.67

	GMM New		GMM New Mis-specified		GMM Mis-specified IV 4 lags+	
	Beta	SD Beta	Beta	SD Beta	Beta	SD Beta
α_1	0.497	0.034	0.379	0.027	-0.211	0.136
α_2	-0.502	0.029	NA	NA	NA	NA
β	0.996	0.05	0.909	0.059	0.926	0.072
First order serial correlation	1	NA	0.984	NA	0.462	NA
Second order serial correlation	0.074	NA	1	NA	1	NA

Table 13: True α_1 = 0.4, α_2 = -0.4, β = 1, $\lambda_{1,2}$ = 0.2 ± 0.6i, modulus = 0.63, periodicity = 4.53

	GMM New		GMM New Mis-specified		GMM Mis-specified IV 4 lags+	
	Beta	SD Beta	Beta	SD Beta	Beta	SD Beta
α_1	0.395	0.039	0.336	0.031	-0.176	0.148
α_2	-0.403	0.031	NA	NA	NA	NA
β	0.997	0.051	0.914	0.055	0.969	0.064
First order serial correlation	1	NA	1	NA	0.304	NA
Second order serial correlation	0.09	NA	1	NA	1	NA

In Tables 11 through 13, going to the appropriate, though weak, instrument makes the sign on the estimated coefficient on Y_{t-1} on the mis-specified but correctly instrumented equation become negative. While that might, at first, appear to be a consequence of weak instrumenting, there is another possible interpretation, which if correct would be consistent with the general thrust of our analysis.

Taking at the cases with complex roots as a group, in the cases where the periodicity was relatively long the correctly instrumented GMM estimation yielded a positive fraction as the coefficient on Y_{t-1}. This was the case, for example, when the cycles had period greater than 10, meaning that the periodicity exceeded the value of T in our panels, and in the case of cycle of 6, which was very long relative to our panel T. On the other hand, for periodicity of roughly 5, the preferred instrumented GMM coefficient in the mis-specified equation dropped virtually to zero and as the periodicity of the cycles (remembering that these are all stable cycles, not permanent cycles, as would be the case if we had a complex unit root) dropped below 5 we began to get negative values for the estimated FODE.

In a FODE, a negative coefficient on the only lagged value of Y produces alternations - a pattern of discrete jumps from above to below the trend of the trajectory. Since a FODE cannot yield true cyclical behavior - that would require complex roots, which in turn would require working with at least a SODE - stable alternations are as close as the estimation procedure can come to tracking a variable whose true path is a stable cycle. Thus a result which may be taken to be the consequence of the instrumenting problem may in fact be a consequence of the omitted variable bias, with the estimation procedure coming as close as it can to tracking the true dynamic process subject to the constraint that it cannot use Y_{t-2} or, to look at it slightly differently, subject to the constraint that the coefficient on Y_{t-2} be set equal to zero.

In essence, what is happening in these complex root cases is a reflection of what the estimation procedure can see. With periodicity of 10, there is a good chance that for many of our individual panels the data will appear monotonic: all lying on the downswing or all on the upswing and only a subset of the individual panels incorporating a cyclical turning point. On the other hand, when the periodicity of the cycle is of the order of 4, all of the individual panels will include turning points, meaning that all of the individual panels will include some observations which are on the downswing and some which are on the upswing. Because we are, by mis-specifying the estimated equation, constraining the coefficient on Y_{t-2} to equal zero, the only way the procedure can fit an equation to the stable (i.e. converging) oscillations it is seeing in the data is with stable alternations, which translate into making the coefficient on Y_{t-1} negative and less than one in absolute value.

VII Conclusions

In this paper we have discussed issues arising from dynamic misspecifications of micro-level estimating equations. We have argued that, while first order dynamics are probably quite adequate for macro and market level analysis, at which dynamics are likely to be adjustment, or transition dynamics, when we work at the level of the individual, as is very often the case in health economics, we need to consider whether we are dealing with intrinsic dynamics, in which case a second order difference equation in the dependent variable would be the appropriate form.

We have argued that, when the latter case applies, we face three intermingled estimation problems. First, we are assuming that the researcher has mis-specified the dynamics of the problem as FODE rather than the true SODE. This is the initial omitted variable problem.

Second, we have Nickell's (1981) problem: we want to estimate the equation using individual level panel data, but each individual in the panel can have a unique intercept, and this can bias the estimate of the coefficient on the lags - a problem which would arise even if the dynamic structure had been specified properly.

Third, because we are dealing with dynamic panel data, we difference to resolve the Nickell problem, and use AB-type GMM estimation to estimate the coefficients of the presumed equation. The residual term of the GMM estimation problem has two issues: one is the omitted variable factor, since the effect of Y_{t-2} is now included in the residuals, the other is an erroneous instrument problem since, having differenced the equation for estimation purposes, we select a lag of Y to use as an instrument based on what we think the dynamic structure is, not on what it actually is. Thus, we do not choose a high enough lag for the instrument and introduce another source of endogeneity bias.

The results of our Monte Carlo experiments suggest a tendency for the coefficient on the one included lag in the mis-specified equation to settle on a value fairly near the value of the larger of the two roots in the true DGP, in the case where the roots are both real, and the dynamics are monotonic and stable. Thus, when a stable SODE is mis-specified as a FODE, there seems to be a tendency for the best fit FODE equation to coefficient on Y_{t-1} which is close to the stable root of the true DGP. In the cases in which the true coefficient on the first lag was larger than one, the estimated coefficient on the one included lag in the mis-specified equation was less than one, and close to the larger of the two roots, meaning that the mis-specification preserved the overall dynamic stability of the DGP, rather than finding a value for the biased coefficient which was close to the true, which would have suggested instability. It seems that, when the dynamics are mis-specified and the researcher is trying to use a FODE to explain a system which is really driven by a SODE, the regression routine is not so much producing biased estimates of the coefficient on the included lag as it is finding coefficients which best let the FODE approximate the dynamics of the true, SODE. Hence, the tendency of the coefficient on the first (and, in the mis-specified equation, only) lag of the dependent variable to settle in to a value close to the value of the largest of the two roots of the true DGP. Thus, in the case where the coefficient on the first lag of Y in the true, SODE DGP, is 1.5 and that on the second lag is -0.56, giving roots of 0.8 and 0.7, the mis-specified FODE form has a coefficient (in Table 1) of 0.796. Clearly this is not equal to the true coefficient on Y_{t-1} but equally clearly it would not be desirable for the estimation procedure to yield a coefficient close to 1.5 in a FODE, since that would imply unstable dynamics whereas the true dynamics are stable. The estimation procedure appears to be matching the dynamics of the true data as well as it can, given that we have restricted it to fitting a FODE.

Our experiments with complex roots[6] suggest that the use of an improper lag of the dependent variable instrument does have an effect on the results. We found there that, when we switched to a more appropriate, though weaker instrument, it seemed that the AB estimation was able to spot the cyclical dynamics so long as the periodicity of the cycle was sufficiently short. This may mean that if strange results like negative signs arise in practical applications, they should be taken as warnings of possible dynamic misspecification. On the other hand, the MC Standard Deviation of those negative coefficients was large relative to the size of the coefficient, so it is not clear how good a warning sign this might be. We also found mixed results with regard to whether the AB m-statistics provide warning of the issue. We should note that, since the m-tests were designed to detect certain types of structure in the disturbance terms, we would not necessarily expect them to provide warning of mis-specified dynamics: our interest in these statistics was strictly by analogy with the macro-econometric literature on mis-specified dynamics, where the Durbin-Watson statistic did quite a creditable job of warning of the problem.

Our results suggest that, given mis-specification of the estimating equation, the biased results are a better approximation to the true dynamics of the system than would be the case if the coefficient on the sole included lagged dependent variable were closer to the true value of α_1. Clearly it would be far preferable to estimate a correctly specified equation, but this result suggests interesting questions about the nature of the bias associated with dynamic mis-specification.

This issue is, we would argue, of interest to health economists for a couple of reasons. One is that, since many of the standard models are models of individual level inter-temporal optimization, SODEs are much more likely to appear in health economics applications. The second is that the longitudinal data sets which are available are, increasingly long enough for estimation of SODEs. The essential dynamic panel data problem, as set out by Nickell (1981), remains when the true DGP is a SODE, and if AB-type GMM estimation is used where an inappropriate lag of the dependent variable is used as the instrument - too short a lag- endogeneity bias will be introduced into the results, as a result of the untreated omitted variable bias. Part of the solution might involve testing down in the dynamics: when we ran experiments (not reported here) in which we estimated

[6]We repeat that cases with complex roots are unlikely to appear in real-world applications: we include them in this paper to give a sense of what is going on.

a SODE when the true DGP was a FODE, the results indicated that, despite the fact that the FODE structure introduces some persistence into the behavior of the lagged dependent variables, the redundant second lag would in general not be significantly different from zero. Finding such a result would then allow for a reduction in the number of lagged dependent variables from 2 to 1 and move to a less weak instrument.

Our results suggest that, in empirical applications where the dynamics are intrinsic rather than a consequence of lagged adjustment, we should devote more attention to investigating the length of the lag than has typically been the case in the literature. Further, our results suggest that while in the case of mis-specified dynamics there is likely to be an endogeneity problem in the instrumenting, there is method to the madness with which the AB type GMM approach estimates the coefficient on the lagged dependent variable: rather than simply being biased and hence simply wrong, that coefficient may actually be informative about the underlying, intrinsic dynamics of the time series data.

References

Arellano, Manuel (2003): *Panel Data Econometrics* Oxford University Press.

Arellano, M. and Bond, S. (1991): "Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations" *Review of Economic Studies* 58, 227-297

Baumol, William J. (1970): *Economic Dynamics* (ed. edn.) Macmillan, London.

Choi, In (2015): "Panel Cointegration" Chapter 2 in Badi Baltagi (ed) The Oxford Handbook of Panel Data Oxford University Press.

Croissant Y, Millo G (2008). "Panel data econometrics in R: The plm package." *Journal of Statistical Software*, 27(2), 1-43.

Ferguson, Brian S. and G. C. Lim (2003): Discrete Time Dynamic Economic Models: Theory and Empirical Applications, Routledge.

Im, Kyung So, M.Hashem Pesaran, Yongcheol Shin (2003): "Testing for unit roots in heterogeneous panels" *Journal of Econometrics*, Volume 115, Issue 1, July 2003, Pages 53-74.

Jung, Hosung (2005): *A Test for Autocorrelation in Dynamic Panel Data Models* Hitotsubashi University Research Unit for Statistical Analysis in Social Sciences, Discussion Paper 77, February

Kruiniger, Hugo (2009): "GMM Estimation and Inference in Dynamic Panel Data Models With Persistent Data" *Econometric Theory* 25, 1348-1391.

Lee, Yoonseok (2012): "Bias in dynamic panel models under time series mis-specification" *Journal of Econometrics* 169, 54-60

Nickell, S. (1981): "Bias in dynamic models with fixed effects" *Econometrica* 49, 1417-1421.

Pesaran, M. Hashem (2015): *Time Series and Panel Data Econometrics* Oxford University Press.

Pesaran, M. Hashem and Ron Smith (1995): "Estimating long-run relationships from dynamic heterogeneous panels" *Journal of Econometrics* 68(1), 79-113.

Phillips, Peter C. B. (2014) "Dynamic Panel GMM With Roots Near Unity", Cowles Foundation Discussion Paper No. 1962, December.

R Core Team (2012). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL `http://www.R-project.org/`

Westerlund, Joakim (2007): "Testing for Error Correction in Panel Data" *Oxford Bulletin of Economics and Statistics*, 69 (6) 709-748.