UvA ECONOMETRICS

Discussion Paper: 2015/03

A Simple Estimator for Short Panels with Common Factors

Artūras Juodis and Vasilis Sarafidis

www.ase.uva.nl/uva-econometrics

Amsterdam School of Economics
Roetersstraat 11
1018 WB AMSTERDAM
The Netherlands

UvA UNIVERSITEIT VAN AMSTERDAM

A Simple Estimator for Short Panels with Common Factors[☆]

Artūras Juodis[a], Vasilis Sarafidis[b,*]

[a] *Amsterdam School of Economics, University of Amsterdam and Tinbergen Institute*
[b] *Department of Econometrics and Business Statistics, Monash University.*

Abstract

There is a substantial theoretical literature on the estimation of short panel data models with common factors nowadays. Nevertheless, such advances appear to have remained largely unnoticed by empirical practitioners. A major reason for this casual observation might be that existing approaches are computationally burdensome and difficult to program. This paper puts forward a simple methodology for estimating panels with multiple factors based on the method of moments approach. The underlying idea involves substituting the unobserved factors with time-specific weighted averages of the variables included in the model. The estimation procedure is easy to implement because unobserved variables are superseded with observed data. Furthermore, since the model is effectively parameterized in a more parsimonious way, the resulting estimator can be asymptotically more efficient than existing ones. Notably, our methodology can easily accommodate observed common factors and unbalanced panels, both of which are important empirical scenarios. We apply our approach to a data set involving a large panel of 4,500 households in New South Wales (Australia), and estimate the price elasticity of urban water demand.

Keywords: Dynamic Panel Data, Factor Model, Fixed T Consistency, Monte Carlo Simulation, Urban Water Management.
JEL: C13, C15, C23.

[☆] We would like to thank participants at IAAE (Thessaloniki, 2015), the Workshop on Panel Data (Amsterdam, 2015), and the Tinbergen Institute. We express special thanks to Pavel Čížek, Rutger Teulings and Frank Windmeijer for helpful comments. Part of this paper was written while the first author enjoyed the hospitality of the Department of Econometrics and Business Statistics at Monash University. Financial support from the NWO MaGW grant "Likelihood-based inference in dynamic panel data models with endogenous covariates" is gratefully acknowledged by the first author.

[*] Corresponding author. 900 Dandenong Road, Caulfield East, Victoria 3145, Australia. E.mail: vasilis.sarafidis@monash.edu

October 30, 2015

1. Introduction

There is an increasingly growing theoretical literature on estimating panel data models with common factors and fixed T, the number of time series observations. The common factor approach is appealing because it allows for several sources of *multiplicative* unobserved heterogeneity and generalizes the highly popular two way error components structure, which represents *additive* heterogeneity. To illustrate, consider the following model:

$$y_{i,t} = \boldsymbol{x}_{i,t}'\boldsymbol{\beta} + u_{i,t}, \quad u_{i,t} = \boldsymbol{\lambda}_i'\boldsymbol{f}_t + \varepsilon_{i,t}; \; i = 1, \ldots, N \quad t = 1, \ldots, T, \tag{1}$$

where both $\boldsymbol{f}_t$ (factors) and $\boldsymbol{\lambda}_i$ (factor loadings) are L dimensional vectors. As an example, consider the study area of analyzing returns to education, or earnings and inequality; the individual wage rate, $y_{i,t}$, is modeled as a function of variables such as education, experience, tenure, gender and race; however, wages may also depend on individual-specific characteristics that are unobserved and typically difficult to measure, like innate ability, skills, and so forth. The two way error components model imposes the restriction that these characteristics are constant over time, while any time-varying unobserved effects are assumed to be common across all individuals. It is easy to see that this formulation is actually a special case of the common factor approach, arising by setting $L = 2$, $\boldsymbol{\lambda}_i = (\gamma_i, 1)'$ and $\boldsymbol{f}_t = (1, \phi_t)'$. One advantage of the multi-factor approach is that it allows the impact of individual-specific unobserved characteristics to vary over time in an intertemporarily arbitrary way. For instance, in the present example one can think of the factors as representing prices for a set of individual-specific skills, which are likely to fluctuate over time according to (say) the business cycle of the economy. By contrast, the two way error components model assumes that prices do not vary over time, which may not be realistic in practice.

The common factor approach is also appealing because it accommodates for time-varying unobserved effects that hit all individual entities, albeit with different intensities. For instance, in the empirical growth literature $\boldsymbol{f}_t$ may represent common shocks, or different streams of time-varying technology, while $\boldsymbol{\lambda}_i$ denotes the rate at which country i absorbs such advances that may be potentially available to all economies; see Bun and Sarafidis (2015) for a detailed discussion on these issues. In the context of estimation of production functions, a subset of the factor component can be viewed as representing different sources of technical efficiency that varies over time, while other factors can capture industry-wide regulatory changes.

In all examples described above, it is not hard to think of reasons why the factor component is most likely to be correlated with the regressors, which renders estimators that accommodate only for a two way error components structure inconsistent.

Perhaps one of the earliest methodologies put forward in dynamic panels to deal with this type of multiplicative heterogeneity was proposed by Holtz-Eakin et al. (1988), and it was subsequently extended by Nauges and Thomas (2003a); this procedure eliminates the common factors using some form of quasi-differencing and utilizes a method of moments approach to estimate the unknown parameters of the model. Ahn et al. (2001) and Ahn et al. (2013) suggested a different way of eliminating the factors, based on quasi-long-differences.

More recently, Robertson and Sarafidis (2015) proposed a GMM approach that introduces new parameters to represent the unobserved covariances between the factor component of the error term and the instruments; in addition, the authors showed that if the structure of the model is autoregressive, there are restrictions in the nuisance parameters that lead to a more efficient GMM estimator compared to existing quasi-differencing approaches. Following Bai (2013), Hayakawa (2012) analyzes a GMM estimator that approximates the factor loadings using a Chamberlain (1982) type projection approach. On the other hand, Bai (2013) proposes a maximum likelihood estimator. Most recently, Hayakawa et al. (2014) propose a related maximum likelihood estimator that eliminates the individual-specific, time-invariant effects from the model using first-differencing and subsequently treats the remaining, genuine factor component as random and uncorrelated with the covariates.

All aforementioned methods are highly non-linear, essentially because the unknown factor component enters into the model in a multiplicative way. As a result, the computational burden involved in implementing these approaches is non-negligible, as it is the case with many non-linear algorithms. Furthermore, these methods can be difficult to program and require starting values for a set of nuisance parameters, for which there is often no underlying theory to suggest any. Local minima-related problems might also arise, particularly for GMM estimators that rely on quasi-differencing.[1]

In this paper we propose a simple alternative methodology that involves substituting the unknown factors with time-specific weighted averages of the variables included in the model. The remaining parameters are estimated using a method of moments estimator, or non-linear least squares. The gains of such strategy are threefold. First, given that the model is effectively parameterized in a more parsimonious way, the resulting estimators

[1] See Kruiniger (2008) on this issue, as well as Juodis and Sarafidis (2014) for a more detailed discussion.

can be asymptotically more efficient than existing ones. Second, the resulting estimation procedure is substantially easier to program because the unobserved factors are superseded by observed data. As a result, the issue of how to initialize nuisance parameters reduces essentially to the rather more trivial task of implementing a grid search of structural parameters. Finally, the simplified moment conditions can be linearized in a straightforward way, leading to a complete linear reparameterization of the model.[2] We remark that the proposed methodology is easily extendable to allow common factors that are observed, and it can accommodate unbalanced panels as well.

Our strategy intuitively resembles the approach employed by the Common Correlated Effects (CCE) estimator proposed by Pesaran (2006), and is also related to the recent work by Karabıyık, Urbain, and Westerlund (2014), who extend the CCE approach using external instruments, in addition to simple cross-sectional averages of the data. The main difference is that the aforementioned papers consider estimation of static panels with strictly exogenous regressors and N, T both large, whereas here the focus is on panels with T fixed, while endogenous regressors are allowed.

The finite sample performance of the proposed estimators is investigated using simulated data. The results indicate that these estimators perform well under a wide range of specifications. The method is applied on a large sample of households in New South Wales (Australia), each one observed over a period of 5 years, in order to estimate the price elasticity of water usage demand. This is an important empirical topic and a subject of ongoing research, not only among economists but also across international environmental agencies, regulators, water utilities and the general public.

The outline of the rest of the paper is as follows. The next section provides some background of the literature and introduces the main idea behind our approach using an autoregressive model with a single factor component. Section 3 extends this to the dynamic panel data model with covariates and multi-factor residuals. Section 4 investigates the finite sample performance of the estimator, and Section 5 applies our approach to provide new evidence on the effect of price elasticity on the demand for water. A final section concludes.

In what follows we briefly introduce our notation. The usual $\text{vec}(\cdot)$ operator denotes the column stacking operator, while $\text{vech}(\cdot)$ is the corresponding operator that stacks only the elements on and below the main diagonal. Shorthand notation $\boldsymbol{x}_{i,s:q}, s \leq q$ is used to denote the vectors of the form $\boldsymbol{x}_{i,s:q} = (x_{i,s}, \dots, x_{i,q})'$. For further details regarding the

[2] Hayakawa (2012) also discusses linearization of the moment conditions proposed by Ahn et al. (2013).

notation used in this paper see Abadir and Magnus (2002).

2. Background and Main Idea

To facilitate exposition of the main ideas of this paper and how these relate to existing work in the panel data GMM literature, this section considers a panel autoregressive model of order one with a single factor. This particular choice is also motivated from the fact that the AR(1) model has become the workhorse for theoretical analysis of panels with weakly exogenous regressors (see e.g. Arellano (2003, Ch. 6)). The more general model with covariates and multiple factors is developed in the next section.

2.1. Basic Model and Existing Literature

Let

$$y_{i,t} = \alpha y_{i,t-1} + u_{i,t}, \quad u_{i,t} = \lambda_i f_t + \varepsilon_{i,t}; \; i = 1, \ldots, N, \quad t = 1, \ldots, T, \tag{2}$$

where $y_{i,t}$ is the observation on the variable of interest for individual unit i at time t, f_t denotes the value of the unobserved factor at time t, λ_i is the individual-specific factor loading and $\varepsilon_{i,t}$ is a purely idiosyncratic error component.

There is a growing literature dealing with the model above when N is large and T fixed. In this case, f_t is typically treated as a fixed parameter to be estimated, while λ_i is stochastic. Holtz-Eakin et al. (1988) and Nauges and Thomas (2003a) propose a GMM estimator based on quasi-differences. In particular, multiplying the lagged value of (2) by $r_t \equiv f_t / f_{t-1}$ and subtracting from the original model yields

$$y_{i,t} - r_t y_{i,t-1} = \alpha \left(y_{i,t-1} - r_t y_{i,t-2} \right) + \varepsilon_{i,t} - r_t \varepsilon_{i,t-1}; \quad t = 2, \ldots, T. \tag{3}$$

Hence the transformed error of the model is free from the factor component because

$$\lambda_i f_t - r_t \lambda_i f_{t-1} = \lambda_i \left(f_t - r_t f_{t-1} \right) = 0. \tag{4}$$

The non-linear moment conditions are of the form

$$\mathrm{E} \left[y_{i,s} \left(\varepsilon_{i,t} - r_t \varepsilon_{i,t-1} \right) \right] = 0; \quad s < t - 1. \tag{5}$$

Ahn et al. (2001) suggest transforming the model to eliminate the factor component in a different way. In particular, normalizing $f_T = 1$ and multiplying the model at period T by f_t yields

$$f_t y_{i,T} = \alpha f_t y_{i,T-1} + \lambda_i f_t + f_t \varepsilon_{i,T}. \tag{6}$$

Subtracting the above equation from (2) yields

$$(y_{i,t} - f_t y_{i,T}) = \alpha \left(y_{i,t-1} - f_t y_{i,T-1}\right) + \left(\varepsilon_{i,t} - f_t \varepsilon_{i,T}\right); \quad t = 1, \ldots, T-1, \tag{7}$$

which can be thought of as an equation expressed in quasi-long-differences. The non-linear moment conditions are of the form

$$\mathrm{E}\left[y_{i,s}\left(\varepsilon_{i,t} - f_t \varepsilon_{i,T}\right)\right] = 0; s < t. \tag{8}$$

As pointed out by Kruiniger (2008), the method by Ahn et al. (2001) requires f_T to be bounded away from zero, or otherwise the quasi-long-differencing transformation described above may fail to converge or be subject to local minima. Similarly, the transformation by Holtz-Eakin et al. (1988) and Nauges and Thomas (2003a) requires that $f_t \neq 0$ for all $t = 1, \ldots, T-1$, in order for r_t to be well-defined in all periods.[3]

Instead of eliminating the factor component by some form of quasi-differencing, Robertson and Sarafidis (2015) propose a GMM estimator that employs the following estimating equations:

$$m_{s,t} = \alpha m_{s,t-1} + g_s f_t; \quad s \leq t,\ t = 1, \ldots, T, \tag{9}$$

where $m_{s,t} = \mathrm{E}\left(y_{i,s} y_{i,t}\right)$ and $g_s = \mathrm{E}\left(y_{i,s}\lambda_i\right)$. Furthermore, they show that a more efficient estimator can be obtained by making use of the fact that, due to the autoregressive nature of the model, the g's have an autoregressive structure as well, which implies a more parsimonious parametrization of the model compared to (9) (their so-called FIVU estimator). To see this, multiply (2) by λ_i and take expectations, which, assuming $\mathrm{E}\left(\lambda_i \varepsilon_{i,t}\right) = 0$, yields

$$g_s = \alpha g_{s-1} + \sigma_\lambda^2 f_s; \quad s = 1, \ldots, T, \tag{10}$$

where $\sigma_\lambda^2 = \mathrm{E}\left(\lambda_i^2\right)$. Thus, knowledge of two extra parameters, σ_λ^2 and g_0, provides a closed form solution for all remaining g's.

Recently, Hayakawa (2012) analyzes a GMM estimator that approximates the factor loadings using a Chamberlain (1982) type projection approach. As shown by Juodis and Sarafidis (2014), the resulting estimator is equivalent to the FIVU estimator that employs the estimating equations in (9).[4]

[3] Juodis and Sarafidis (2014) discuss these issues further and investigate the finite sample properties of several dynamic panel estimators using simulated data.

[4] We do not discuss the estimators by Bai (2013) and Hayakawa et al. (2014) because they both rely on maximum likelihood and require strictly exogenous covariates, which is an assumption that is violated in our empirical application.

The aforementioned estimators are non-linear and can be difficult to program. In addition, they require starting values for a large set of unknown *nuisance* parameters. For instance, the GMM estimator that makes use solely of (9) without further restrictions requires starting values either for $\{f_1, \ldots, f_T\}$ or $\{g_0, \ldots, g_{T-1}\}$. Naturally, multiple local minima-related issues might also arise in this case, as it is true for the estimators involving some form of quasi-differencing.

2.2. Main Idea

The key feature of our approach is to replace f_t with time-specific weighted averages of observed data. Thus, in this respect our strategy resembles intuitively the one employed by the CCE estimator proposed by Pesaran (2006), which is valid for panels with exogenous regressors and N, T both large.

To illustrate this, let w_i denote a weight for individual i that satisfies $N^{-1}\sum_{i=1}^{N} w_i \varepsilon_{i,t} \xrightarrow{p} \mathrm{E}\left(w_i \varepsilon_{i,t}\right) = 0$ as $N \to \infty$. Define $m_t \equiv \mathrm{E}\left(w_i y_{i,t}\right)$. Thus, multiplying (2) by w_i and taking expectations yields

$$m_t = \alpha m_{t-1} + \mu_\lambda f_t;\ t = 1, \ldots, T, \tag{11}$$

where $\mu_\lambda = \mathrm{E}\left(w_i \lambda_i\right) \neq 0$ (by assumption). Hence, solving for f_t yields

$$f_t = \left(m_t - \alpha m_{t-1}\right)/\mu_\lambda. \tag{12}$$

Within the GMM framework, the assumption $\mu_\lambda \neq 0$ is actually easy to verify, as it will be shown. Replacing f_t in the original model (2) by the expression above we obtain

$$y_{i,t} = \alpha y_{i,t-1} + \upsilon_{i,t}; \quad \upsilon_{i,t} = \lambda_i \left(m_t - \alpha m_{t-1}\right) + \varepsilon_{i,t}, \tag{13}$$

where, to save on notation, λ_i is reparametrized in terms of the original factor loading divided by μ_λ, i.e. $\lambda_i \equiv \lambda_i/\mu_\lambda$. Thus, multiplying (13) by $y_{i,s}$ and taking expectations gives rise to the following estimating equations

$$m_{s,t} = \alpha m_{s,t-1} + g_s \left(m_t - \alpha m_{t-1}\right); \quad s < t,\, t = 1, \ldots, T. \tag{14}$$

The "structural" parameter, α, and the nuisance parameters (the g's) can be estimated using a straightforward application of non-linear least squares in (14), or a method of moments estimator that is based on

$$\mathrm{E}\left[y_{i,s}\upsilon_{i,t} - g_s \left(m_t - \alpha m_{t-1}\right)\right] = 0; \quad s < t. \tag{15}$$

Notably, starting values for the g's can be obtained very easily because choosing a specific value for α provides a closed form solution for g_s, as it can be seen from (14); thus, the optimal solution can be obtained using an iterative procedure, where given α we estimate g_s, and given g_s we estimate α. Subsequently, a grid search for α can be implemented such that $\hat{\alpha}$ is chosen as the value of the autoregressive parameter that corresponds to the smallest value of the objective function.

The estimation problem can even be linearized, at the expense of introducing extra parameters. In particular, expanding the last term in (14) and setting $g_s^* \equiv \alpha g_s$ means that α can be estimated using standard least squares. The model is exactly identified for $T = 4$ and overidentified for $T > 4$. The identifying assumption is, again, $\mu_\lambda \neq 0$.[5]

There are two important issues that emerge following the analysis above. Firstly, the choice of w_i used to obtain (11); and secondly, how to verify the assumption $\mu_\lambda \neq 0$ such that cross-sectional averages of the data can span f_t.

2.3. Choices for w_i

One choice for the weights is to simply set $w_i = 1$, such that m_t in (11) becomes the crude time-specific average of $y_{i,t}$. This particular strategy resembles the CCE estimator of Pesaran (2006), as in this case the model is augmented with cross-sectional averages of observed variables. The choice requires that $\mathrm{E}(\lambda_i) \neq 0$, which can be a plausible condition in several applications. In their Monte Carlo study, Juodis and Sarafidis (2014) show that estimators that rely on normalizing factor-specific values (e.g. as in Ahn et al. (2013)) can be sensitive to the underlying DGP for f_t. Hence linearizing the model with respect to non-stochastic weights $w_i = 1$ may result in an estimate with superior properties, assuming $\mathrm{E}(\lambda_i) \neq 0$.

Remark 1. This particular choice for w_i in the present model yields an estimator that is asymptotically more efficient than the FIVU estimator proposed by Robertson and Sarafidis (2015), which employs the estimating equations given in (9). This is because the former requires estimating $T + 1$ parameters, while the latter estimates $1 + 2T - 1$ parameters. To put it differently, the former estimator utilizes T additional moment conditions, i.e. those in equation (11). Of course, these become redundant when $\mu_\lambda = 0$. The resulting estimator is also asymptotically more efficient than the GMM estimators proposed by Ahn et al. (2001) and Nauges and Thomas (2003a) for exactly the same reason: it makes use of additional information, which results in a more parsimonious parameterization of the model.

[5]Clearly, the linearized estimator can provide an extra starting value of α for the iterative procedure corresponding to (15).

An alternative strategy is to choose weights with respect to the observed data. As an example, one may set $w_i = y_{i,0}$, which implies that (11) can be expressed as

$$m_{0,t} = \alpha m_{0,t-1} + g_0 f_t. \tag{16}$$

Notice that this requires $g_0 \neq 0$, which will hold true unless the initial observation is uncorrelated with λ_i, an implausible condition that violates the "fixed effects framework".[6]

Remark 2. Setting $w_i = y_{i,0}$ implies that the T moment conditions with respect to $y_{i,0}$ have been "consumed" to approximate f_t. Therefore, the estimating equations in (14) apply for $s > 0$ and $t > 1$. In other words, T moment conditions are effectively dropped out in order to solve in terms of T unknown parameters, the f's. Thus, one could view this estimator as a "normalized" version of FIVU, in which the estimating equations are divided by g_0.

Alternative choices of w_i can be obtained based on $y_{i,s}$, $s > 0$, or on powers of $y_{i,0}$, such as $y_{i,0}^2$ and so on. Although the latter strategy might appear to require non-zero third moments, it is worth emphasizing here that since we employ non-central moments, these can be different from zero even if the distribution of $y_{i,t}$ is symmetric. Furthermore, in practice it is easy to verify whether this is true or not because for (say) $w_i = y_{i,0}^2$, $\mathrm{E}\left(y_{i,0}^2 y_{i,t}\right)$ can be trivially estimated using the sample average of $w_i y_{i,t}$.

Note that the approach developed in this paper is related to the recent work by Karabıyık, Urbain, and Westerlund (2014), who suggest using external instruments instead of simple cross-sectional averages and build an extended Common Correlated Estimator (CCE) of Pesaran (2006). Given that external instruments can also be used within our approach, the major difference is that the aforementioned paper considers consistent estimation of static panels with strictly exogenous regressors for large N and T, whereas here the focus is on panels with T fixed, and endogeneity is allowed.

2.4. Verifying the condition $\mu_\lambda \neq 0$

The condition $\mu_\lambda \neq 0$ is verifiable in two different ways. First of all, it is straightforward to see that $\mu_\lambda = 0$ implies that the factor is not spanned by weighted averages of the observed data. Therefore, the model becomes mis-specified and thereby the moment

[6] Furthermore, the autoregressive nature of the model suggests that the initial condition $y_{i,0}$ is a weighted average of the pre-sample $\{\boldsymbol{f}_t\}_{t=0}^{-S}$. Thus even if one $\boldsymbol{f}_t \approx 0$ the total contribution of factors need not be negligible.

conditions are violated. This violation will be reflected in the overidentifying restrictions test statistic that is readily available within our procedure.

Secondly, the condition can be verified based on an examination of the rank of the matrix that contains the observed data. To see this, notice that for $\mu_\lambda = 0$ m_t in (11) becomes just a multiple scalar of m_{t-1}. As a result, collecting these moments in the $[T \times 2]$ matrix $\boldsymbol{H} = (\boldsymbol{m}, \boldsymbol{m}_{-1})$, where $\boldsymbol{m} = (m_1, \ldots, m_T)'$ and $\boldsymbol{m}_{-1} = (m_0, \ldots, m_{T-1})'$, $\mu_\lambda = 0$ implies that $rank\,(\boldsymbol{H}) = 1$, i.e. $\boldsymbol{H}$ is rank-deficient. A similar result applies to models with $L > 1$.

Remark 3. The condition $\mu_\lambda \neq 0$ is effectively analogous to the assumption that $f_T \neq 0$ in Ahn et al. (2013) or to the rank condition listed in equation (21) in Pesaran (2006). The aforementioned papers do not provide a way to be able to verify in practice these assumptions, which, it is fair to say, are taken for granted by empirical practitioners.

2.5. Basic Model with Observed Factors

A useful extension that is important to consider is a model with additional *observed* factors. This is particularly relevant in a large number of applications, where some variables are common across individuals (especially those measured at the macro level, such as interest rates, inflation rate, unemployment rate etc.). Therefore, these variables can be viewed as common *observed* factors.

To illustrate, let

$$y_{i,t} = \alpha y_{i,t-1} + \beta_i t + \lambda_i f_t + \varepsilon_{i,t}; \quad i = 1, \ldots, N,\, t = 1, \ldots, T, \tag{17}$$

so that the model contains two factors, but one of them represents a known deterministic trend. This particular choice is to simplify notation, without loss of generality. Taking expectations yields

$$m_t = \alpha m_{t-1} + \mu_\beta t + \mu_\lambda f_t, \tag{18}$$

where $\mu_\beta = \mathrm{E}\,(\beta_i)$, which is not necessarily different from zero. Thus, solving for f_t and replacing its value in the original equation yields

$$y_{i,t} = \alpha y_{i,t-1} + \upsilon_{i,t}; \quad \upsilon_{i,t} = \beta_i t + \lambda_i \left(m_t - \alpha m_{t-1} - \mu_\beta t\right) + \varepsilon_{i,t}, \tag{19}$$

where, to save on notation, λ_i is again defined in terms of the original factor loading divided by μ_λ.

As a result, multiplying the above model by $y_{i,s}$ and taking expectations yields

$$m_{s,t} = \alpha m_{s,t-1} + g_s \left(m_t - \alpha m_{t-1}\right) + \gamma_s t, \tag{20}$$

where $\gamma_s \equiv (\gamma_s^* - \mu_\beta)$, $\gamma_s^* \equiv [\mathrm{E}(y_{i,s}\beta_i) - g_s\mu_\beta]$ and g_s has been defined previously. Effectively, equation (20) merely augments equation (14) by $\gamma_s t$. Thus as before, picking a value for α implies a closed form solution for g_s (and γ_s). Estimation can be implemented using a simple iterative procedure as discussed above.

In conclusion, our approach can accommodate common observed factors in a straightforward way. This is not the case with many of the estimators discussed in Section 2.1 (see Section 4.3. in Juodis and Sarafidis (2014) for a detailed discussion on this issue).

3. General Case

3.1. Model and Assumptions

Motivated by our application, we now consider the following dynamic panel data model with a multi-factor error structure:

$$y_{i,t} = \alpha y_{i,t-1} + \sum_{k=1}^{K} \beta_k x_{i,t}^{(k)} + \boldsymbol{\lambda}_i' \boldsymbol{f}_t + \varepsilon_{i,t}; \quad i = 1, \ldots, \quad N, t = 1, \ldots, T. \tag{21}$$

The dimension of the unobserved components $\boldsymbol{\lambda}_i$ and $\boldsymbol{f}_t$ is $[L \times 1]$. The main parameters of interest are the "structural" parameters $(\alpha, \beta_1, \ldots, \beta_K)'$, which are all bounded by a finite constant. For convenience we stack the observations over time for each individual i so that the model can be rewritten in the following manner:

$$\boldsymbol{y}_i = \alpha \boldsymbol{y}_{i,-1} + \sum_{k=1}^{K} \beta_k \boldsymbol{x}_i^{(k)} + \boldsymbol{F}\boldsymbol{\lambda}_i + \boldsymbol{\varepsilon}_i, \quad i = 1, \ldots, N, \tag{22}$$

where $\boldsymbol{y}_i$ is defined as $\boldsymbol{y}_i = (y_{i,1}, \ldots, y_{i,T})'$ and similarly for $(\boldsymbol{y}_{i,-1}, \boldsymbol{x}_i^{(k)})$, while $\boldsymbol{F} = (\boldsymbol{f}_1, \ldots, \boldsymbol{f}_T)'$ is of dimension $[T \times L]$. In what follows we specify a set of assumptions commonly employed in the literature, followed by some discussion.

Assumption 1: (i) $y_{i,0}$ and $x_{i,t}^{(k)}$ have finite moments up to second order (for all k); (ii) $\varepsilon_{i,t} \sim i.i.d.\,(0, \sigma_\varepsilon^2)$ and has finite moments up to second order; (iii) $\boldsymbol{\lambda}_i \sim i.i.d.\,(\boldsymbol{0}, \boldsymbol{\Sigma}_\lambda)$ with finite moments up to second order, where $\boldsymbol{\Sigma}_\lambda$ is an $[L_0 \times L_0]$ positive definite matrix, and L_0 denotes the true number of factors. $\boldsymbol{F}$ is non-stochastic and uniformly bounded such that $||\boldsymbol{F}|| < b < \infty$.

Assumption 2: $\mathrm{E}\left(\varepsilon_{i,t} | \boldsymbol{y}_{i,0:t-1}, \boldsymbol{\lambda}_i', \boldsymbol{x}_{i,1:\tau(t,1)}^{(1)}, \ldots, \boldsymbol{x}_{i,1:\tau(t,K)}^{(K)}\right) = 0$ for all t, for some positive integers $\tau(t,1), \ldots, \tau(t,K)$.

Assumption 1(i) is a standard regularity condition. Assumptions 1(ii)-1(iii) are employed mainly for simplicity and can be relaxed to some extent.[7] For example, $\varepsilon_{i,t}$ can be heteroskedastic across both dimensions, provided that a sandwich type formula for the variance-covariance matrix of the estimator is used. Conditional moments of $\boldsymbol{\lambda}_i$ can also be heteroskedastic. The independence assumption across i can be relaxed as well, so long as there are sufficient regularity conditions such that probability limits of the terms defined below converge to expectations. We refrain from embarking on such generalizations in order to avoid unnecessary notational cluttering.

Assumption 2 characterises the exogeneity properties of the covariates. In particular, covariates that satisfy $\tau(t,k) = T$ ($\tau(t,k) = t$) are strictly (weakly) exogenous with respect to the idiosyncratic error component, otherwise they are endogenous. The estimator proposed in this paper can allow for strictly/weakly exogenous and endogenous regressors. In addition, Assumption 2 implies that the idiosyncratic errors are conditionally serially uncorrelated. Again, this can be relaxed in a relatively straightforward way; for example, one could assume instead that either $\mathrm{E}\left(\varepsilon_{i,t}|\boldsymbol{y}_{i,0:q}, \boldsymbol{\lambda}_i', \boldsymbol{x}_{i,0:\tau(t,1)}^{(1)}, \ldots, \boldsymbol{x}_{i,0:\tau(t,K)}^{(K)}\right) = 0$, where $q < t-1$, or $\mathrm{E}\left(\varepsilon_{i,t}|\boldsymbol{\lambda}_i', \boldsymbol{x}_{i,0:\tau(t,1)}^{(1)}, \ldots, \boldsymbol{x}_{i,0:\tau(t,K)}^{(K)}\right) = 0$. In the former case a moving average process of a certain order in $\varepsilon_{i,t}$ is permitted and moment conditions with respect to (lagged values of) $y_{i,q}$ can be used. In the latter case, an autoregressive process in $\varepsilon_{i,t}$ is permitted and moment conditions with respect to (lagged values of) $x_{i,\tau(t,k)}^{(k)}$ remain valid. Finally, Assumption 2 implies that the idiosyncratic error is conditionally uncorrelated with the factor loadings. This is required for identification based on internal instruments in levels and it can be relaxed to some extent, at the expense of considerable extra computational burden. Notice lastly that the set of our assumptions implies that $y_{i,t}$ has finite second-order moments, but it does not imply conditional homoskedasticity for the error components.

Suppose that there exists an $[L \times 1]$ vector of individual-specific weights $\boldsymbol{w}_i$ that satisfies

$$N^{-1}\sum_{i}^{N} \boldsymbol{\varepsilon}_i \boldsymbol{w}_i' \xrightarrow{p} \mathrm{E}\left(\boldsymbol{\varepsilon}_i \boldsymbol{w}_i'\right) = \mathbf{O}_{T\times L}.$$

Therefore, post-multiplying (22) by $\boldsymbol{w}_i'$ and taking expectations we obtain

$$\boldsymbol{M}_y = \alpha \boldsymbol{M}_{y,-1} + \sum_{k=1}^{K} \beta_k \boldsymbol{M}_x^{(k)} + \boldsymbol{F}\boldsymbol{G}_w, \tag{23}$$

[7] The zero-mean assumption for $\varepsilon_{i,t}$ is actually implied by Assumption 2.

where $\boldsymbol{M}_y = \mathrm{E}(\boldsymbol{y}_i \boldsymbol{w}_i')$, and so on for the remaining "M" (moments) matrices, while $\boldsymbol{G}_w = \mathrm{E}(\boldsymbol{\lambda}_i \boldsymbol{w}_i')$, is an $[L \times L]$ matrix. Assuming that $\boldsymbol{G}_w$ has full rank, one can solve for $\boldsymbol{F}$ as follows:

$$\boldsymbol{F} = \left(\boldsymbol{M}_y - \alpha \boldsymbol{M}_{y,-1} - \sum_{k=1}^{K} \beta_k \boldsymbol{M}_x^{(k)} \right) \boldsymbol{G}_w^{-1}.$$

Replacing the above value of $\boldsymbol{F}$ in the original model expressed in vector form yields

$$\begin{aligned} \boldsymbol{y}_i &= \alpha \boldsymbol{y}_{i,-1} + \sum_{k=1}^{K} \beta_k \boldsymbol{x}_i^{(k)} + \left(\boldsymbol{M}_y - \alpha \boldsymbol{M}_{y,-1} - \sum_{k=1}^{K} \beta_k \boldsymbol{M}_x^{(k)} \right) \boldsymbol{\lambda}_i + \boldsymbol{\varepsilon}_i \\ &= \alpha \boldsymbol{y}_{i,-1} + \sum_{k=1}^{K} \beta_k \boldsymbol{x}_i^{(k)} + (\boldsymbol{I}_T \otimes \boldsymbol{\lambda}_i') \operatorname{vec} \left(\boldsymbol{M}_y' - \alpha \boldsymbol{M}_{y,-1}' - \sum_{k=1}^{K} \beta_k \left(\boldsymbol{M}_x^{(k)} \right)' \right) + \boldsymbol{\varepsilon}_i, \end{aligned} \tag{24}$$

where, as in the model with one factor, $\boldsymbol{\lambda}_i$ is redefined in terms of $\boldsymbol{G}_w^{-1}$, i.e. $\boldsymbol{\lambda}_i \equiv \boldsymbol{G}_w^{-1} \boldsymbol{\lambda}_i$.

Let $\boldsymbol{z}_i = \left(\boldsymbol{y}_{i,-1}', \left(\boldsymbol{x}_{i,0:\tau(T,1)}^{(1)} \right)', \ldots, \left(\boldsymbol{x}_{i,0:\tau(T,K)}^{(K)} \right)' \right)'$ denote a $[d \times 1]$ vector that contains all available instruments, the size of which depends on the number of variables employed as instruments and their exogeneity properties.[8] Also, let $\boldsymbol{S} = \operatorname{diag}(\boldsymbol{S}_1, \ldots, \boldsymbol{S}_T)$ denote a block diagonal matrix with a typical (block-)diagonal entry equal to $\boldsymbol{S}_t$, where $\boldsymbol{S}_t$ is a $[\zeta_t \times d]$ selection matrix of zeros and ones that picks from the vector $\boldsymbol{z}_i$ ζ_t valid instruments at time t. As a result, $\boldsymbol{S}$ has dimension $[\zeta \times dT]$, where $\zeta = \sum_{t=1}^{T} \zeta_t$. Define $\boldsymbol{Z}_i' \equiv \boldsymbol{S}(\boldsymbol{I}_T \otimes \boldsymbol{z}_i)$. Under Assumptions 1-2, the following set of population moment conditions is valid by construction:

$$\mathrm{E}(\boldsymbol{Z}_i' \boldsymbol{\varepsilon}_i) = \boldsymbol{S}\, \mathrm{E}(\operatorname{vec}(\boldsymbol{z}_i \boldsymbol{\varepsilon}_i')) = \boldsymbol{0}_\zeta. \tag{25}$$

Thus, pre-multiplying (24) by $\boldsymbol{Z}_i'$ and taking expectations yields

$$\begin{aligned} \boldsymbol{m} &= \alpha \boldsymbol{m}_{-1} + \sum_{k=1}^{K} \beta_k \boldsymbol{m}_k + \boldsymbol{S}(\boldsymbol{I}_T \otimes \boldsymbol{G}) \operatorname{vec} \left(\boldsymbol{M}_y' - \alpha \boldsymbol{M}_{y,-1}' - \sum_{k=1}^{K} \beta_k \left(\boldsymbol{M}_x^{(k)} \right)' \right) \\ &= \alpha \boldsymbol{m}_{-1} + \sum_{k=1}^{K} \beta_k \boldsymbol{m}_k + \boldsymbol{S} \operatorname{vec} \left(\boldsymbol{G} \boldsymbol{M}_y' - \alpha \boldsymbol{G} \boldsymbol{M}_{y,-1}' - \sum_{k=1}^{K} \beta_k \boldsymbol{G} \left(\boldsymbol{M}_x^{(k)} \right)' \right), \end{aligned} \tag{26}$$

or

$$\boldsymbol{m} = \alpha \boldsymbol{m}_{-1} + \sum_{k=1}^{K} \beta_k \boldsymbol{m}_k + \boldsymbol{S} \left[\left(\boldsymbol{M}_y - \alpha \boldsymbol{M}_{y,-1} - \sum_{k=1}^{K} \beta_k \boldsymbol{M}_x^{(k)} \right) \otimes \boldsymbol{I}_d \right] \boldsymbol{g}, \tag{27}$$

[8] Here we implicitly assume that for each k, if $q \le t$ then also $\tau(q,1) \le \tau(t,1)$, which is a reasonable assumption so long as $\varepsilon_{i,t}$ is i.i.d.

where $\boldsymbol{g} = \text{vec}(\boldsymbol{G})$ and $\boldsymbol{G} = \text{E}(\boldsymbol{z}_i \boldsymbol{\lambda}_i')$, a matrix that contains nuisance parameters and has dimension $[d \times L]$. Observe that $\boldsymbol{F}$ has been superseded by observed data. As such, starting values for the "structural" parameters provide a close form solution for a set of starting values for $\boldsymbol{g}$. Furthermore, optimization can be performed iteratively in a straightforward way.

As the vector of estimating equations in (27) stands, not all dL nuisance parameters in $\boldsymbol{g}$ can be uniquely identified. The total number of identifiable parameters (up to a normalization) depends on the number of factors (L), as well as on the exogeneity properties of the regressors ($\tau(T,k)$s). In particular, the number of identifiable parameters is given by

$$\#parameters = K + 1 + d \times L - \xi(L, \tau(T,1), \ldots, \tau(T,K)).$$

To illustrate what the $\xi(L, \tau(T,1), \ldots, \tau(T,K))$ function looks like, consider the case where all regressors are weakly exogenous, i.e. $\tau(t,k) = t$ for all k. Notice that the moment conditions with respect to $y_{i,t}$ are of triangular form because $y_{i,t}$ itself and future values of $y_{i,t}$ are not valid instruments, i.e. $E(y_{i,s}\varepsilon_{i,t}) \neq 0$ for $s \geq t$. In the present example, the same holds for the covariates since they are assumed not to be strictly exogenous. As a result, $\boldsymbol{G}$ can be identified only in a block triangular fashion as well. To see this consider equation (21) at period $t = T$:

$$y_{i,T} = \alpha y_{i,T-1} + \sum_{k=1}^{K} \beta_k x_{i,T}^{(k)} + \boldsymbol{\lambda}_i' \boldsymbol{f}_T + \varepsilon_{i,T}.$$

At this time period one can use present/lagged values of $y_{i,T-1}$ and present/lagged values of $x_{i,T}^{(k)}$ as instruments. However, due to the weak exogeneity assumption of all regressors, the following parameters in $\boldsymbol{g}$ appear only in the equation at time $t = T$:

$$\text{E}[y_{i,T-1}\boldsymbol{\lambda}_i'], \text{E}[x_{i,T}^{(1)}\boldsymbol{\lambda}_i'], \ldots, \text{E}[x_{i,T}^{(K)}\boldsymbol{\lambda}_i']. \tag{28}$$

Thus, there are $K+1$ estimating equations that make use of $y_{i,T-1}$ and $x_{i,T}^{(k)}$, $k = 1, \ldots, K$, as instruments and $L(K+1)$ nuisance parameters. Hence, for $L > 1$ one can identify only certain linear combinations of the parameter vector $\boldsymbol{g}$, that is, identification of $\boldsymbol{g}$ is up to a normalization. As a result, following Juodis and Sarafidis (2014) one can conclude that

$$\xi(L, \tau(T,1), \ldots, \tau(T,K)) = (K+1)\frac{L(L-1)}{2}. \tag{29}$$

Remark 4. If one relaxes the weak exogeneity assumption and instead assumes that all regressors are endogenous (i.e. $\tau(t,k) = t - 1$), the form of $\xi(L, \tau(T,1), \ldots, \tau(T,K))$ remains unchanged.

Let $\boldsymbol{\theta} = (\alpha, \beta_1, \ldots, \beta_K, \boldsymbol{g}_r')' \in \Theta$, where $\boldsymbol{g}_r$ denotes the vector of the remaining free parameters in $\boldsymbol{g}$ following a particular set of normalizations, and let Θ denote the full parameter space of $\boldsymbol{\theta}$. Define

$$\boldsymbol{\mu}_i(\boldsymbol{\theta}) = \boldsymbol{Z}_i'\left(\boldsymbol{y}_i - \alpha\boldsymbol{y}_{i,-1} - \sum_{k=1}^{K}\beta_k\boldsymbol{x}_i^{(k)}\right) - \boldsymbol{S}\left[\left(\left(\boldsymbol{y}_i - \alpha\boldsymbol{y}_{i,-1} - \sum_{k=1}^{K}\beta_k\boldsymbol{x}_i^{(k)}\right)\boldsymbol{w}_i'\right)\otimes\boldsymbol{I}_d\right]\widetilde{\boldsymbol{g}}, \tag{30}$$

where $\widetilde{\boldsymbol{g}} = \left(\boldsymbol{g}', (\boldsymbol{g}\backslash\boldsymbol{g}_r)'\right)'$ and $\boldsymbol{g}\backslash\boldsymbol{g}_r$ denotes the part of $\boldsymbol{g}$ not in $\boldsymbol{g}_r$.

Assumption 3: Θ is compact and contains $\boldsymbol{\theta}_0$ in its interior, where $\boldsymbol{\theta}_0$ denotes the true parameter vector. In addition, $\boldsymbol{\theta}_0$ is identified on Θ such that $\mathrm{E}[\boldsymbol{\mu}_i(\boldsymbol{\theta})] = 0$ iff $\boldsymbol{\theta} = \boldsymbol{\theta}_0$.

Assumption 4: $\boldsymbol{\Gamma} \equiv \mathrm{E}[\partial\boldsymbol{\mu}_i(\boldsymbol{\theta})/\partial\boldsymbol{\theta}']_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}$ and $\boldsymbol{\Delta} \equiv \mathrm{E}\left[\boldsymbol{\mu}_i(\boldsymbol{\theta}_0)\boldsymbol{\mu}_i(\boldsymbol{\theta}_0)'\right]$ exist and are full rank matrices.

The following proposition summarizes the asymptotic properties of the proposed estimator.

Proposition 1. *Let* $\boldsymbol{\mu}_N(\boldsymbol{\theta}) = N^{-1}\sum_{i=1}^{N}\boldsymbol{\mu}_i(\boldsymbol{\theta})$ *and define*

$$\widehat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}\in\Theta}\ \boldsymbol{\mu}_N(\boldsymbol{\theta})'\boldsymbol{\Omega}_N\boldsymbol{\mu}_N(\boldsymbol{\theta}),$$

where $\boldsymbol{\Omega}_N$ *is a given positive definite matrix. Then under Assumptions 1-4,* $\hat{\boldsymbol{\theta}}$ *converges in probability to* $\boldsymbol{\theta}_0$ *and*

$$\sqrt{N}\left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right) \xrightarrow{d} N\left(\boldsymbol{0}, (\boldsymbol{\Gamma}'\boldsymbol{\Omega}\boldsymbol{\Gamma})^{-1}(\boldsymbol{\Gamma}'\boldsymbol{\Omega}\boldsymbol{\Delta}\boldsymbol{\Omega}\boldsymbol{\Gamma})(\boldsymbol{\Gamma}'\boldsymbol{\Omega}\boldsymbol{\Gamma})^{-1}\right).$$

Proof. *The proof follows directly from Robertson and Sarafidis (2015).*

Here $\boldsymbol{\Omega} = \mathrm{plim}_{N\to\infty}\boldsymbol{\Omega}_N$. If $\boldsymbol{\Omega}_N = \boldsymbol{\Omega}$ is chosen as $\boldsymbol{\Delta}^{-1}$, then the resulting GMM estimator is optimal in the class of GMM estimators that make use of the moment conditions in (25). The same result applies if one replaces $\boldsymbol{\Delta}$ by a consistent estimate.

Remark 5. As shown by Robertson and Sarafidis (2015), the choice of the identification scheme (i.e. the set of normalizing restrictions) on the vector of nuisance parameters $\boldsymbol{g}$ is not important. Moreover, if one is interested only in estimating the structural parameters of the model, it is not even necessary to impose normalizing restrictions, rather, it suffices that such a normalizing scheme exists.[9] In particular, the GMM estimator of

[9] See Theorem 3 in their paper.

the structural parameters that involves optimizing the objective function with respect to $\boldsymbol{\vartheta} \equiv (\alpha, \beta_1, \ldots, \beta_K, \boldsymbol{g}')'$ will coincide asymptotically with the estimator that optimizes with respect to $\boldsymbol{\theta}$. As a result, the structural parameters of the model can be estimated in practice based on a simple iterative procedure using the estimating equations in (27).

Remark 6. The proposed estimator can be trivially extended to unbalanced panels by simply introducing indicators, as it is the case for the standard fixed effects estimator, depending on whether a particular moment condition is available for individual i or not. This is not the case with many other dynamic panel estimators with common factors, as it is discussed in detail in Section 4.2. in Juodis and Sarafidis (2014).

3.2. A Linearized Version of the Estimator

The vector of estimating equations from the previous section can be easily linearized by defining in (26) the new parameters $\boldsymbol{G}_0 \equiv -\alpha \boldsymbol{G}$, $\boldsymbol{G}_k \equiv -\beta_k \boldsymbol{G}$. In this case the moment conditions become linear. Hence the total number of parameters in (26) is given by

$$\#parameters = K + 1 + dL + d(K+1)L.$$

Compared to (26), its linearized version contains $d(K+1)L$ additional parameters. Essentially, instead of estimating the model with L unobserved factors one can now estimate the model with $\tilde{L} \equiv L(K+2)$ *observed* factors.

Similarly to the discussion regarding identification in the previous section, not all elements of the "$\boldsymbol{G}$" matrices can be uniquely identified in this case. For example, assuming no serial correlation in $\varepsilon_{i,t}$, the number of moment conditions with respect to lagged values of $y_{i,t}$ and the number of the corresponding parameters that are identifiable is given by

$$\#moments(\boldsymbol{y}_{i,-1}) = \frac{T(T+1)}{2}, \quad \#parameters(\boldsymbol{y}_{i,-1}) = 1 + \left(T - \frac{\tilde{L}-1}{2}\right)\tilde{L}.$$

Here as in Section 3.1, the normalization term $\tilde{L}(\tilde{L}-1)/2$ corresponds to the unobserved "last" elements of the $\mathrm{E}[\boldsymbol{y}_{i,-1}\boldsymbol{\lambda}_i']$ that cannot be separately identified, see also Juodis and Sarafidis (2014). Alternatively, without loss in efficiency one can drop the $\tilde{L}(\tilde{L}+1)/2$ lower triangular moment conditions associated with lagged values of $y_{i,t}$ and the same number of corresponding parameters, such that

$$\#moments(\boldsymbol{y}_{i,-1}) = \frac{T(T+1)}{2} - \frac{\tilde{L}(\tilde{L}+1)}{2}, \quad \#parameters(\boldsymbol{y}_{i,-1}) = K + 1 + \left(T - \tilde{L}\right)\tilde{L}. \tag{31}$$

Letting $b = T - \tilde{L}$, the aforementioned *lower triangular moment conditions* are given by

$$\mathrm{E}(\boldsymbol{\varepsilon}_i(\boldsymbol{y}_{i,-1} - \boldsymbol{g}\boldsymbol{w}_i)') = \begin{pmatrix} m_{11} & & & & \\ & \ddots & & & \\ & & m_{b,b} & & \\ & & \vdots & m_{b+1,b+1} & \\ & & & \vdots & \ddots & \\ m_{T,1} & & m_{T,b} & m_{T,b+1} & \cdots & m_{T,T} \end{pmatrix},$$

forming a triangle with corners $(m_{b+1,b+1}, m_{T,T}, m_{T,b+1})$.

Pooling all the moment conditions available with respect to the covariates in this form, one obtains the following number of moment conditions and identifiable parameters:

$$\#moments = \zeta - \frac{\tilde{L}(\tilde{L}+1)}{2}(K+1), \quad \#parameters = (K+1)\left(1 + \left(T - \tilde{L}\right)\tilde{L}\right). \quad (32)$$

3.3. Testing for the rank of $\boldsymbol{G}_w$

As in the simple case, the full rank assumption on $\boldsymbol{G}_w$ can be verified either by using the overidentifying restrictions test statistic, or based on the rank of a matrix that contains observed data. Since the former way is identical to what we have already discussed in Section 2.4, we focus on the latter.

Define the following matrix

$$\boldsymbol{M} \equiv [\boldsymbol{M}_y, \boldsymbol{M}_{y,-1}, \boldsymbol{M}_1, \dots, \boldsymbol{M}_K],$$

which is of dimension $[T \times (K+2)L]$. The following theorem is fundamental for our approach.

Theorem 1. *Suppose that* $\boldsymbol{M}$ *is a full rank matrix, i.e.* $\mathrm{rk}(\boldsymbol{M}) = (K+2)L$. *Then* $\boldsymbol{G}_w$ *is invertible.*

Proof. *The proof is provided in the Appendix.*

As a result of the theorem above, one could in principle verify, prior to estimation, whether the unobserved matrix $\boldsymbol{G}_w$ has full rank by checking if the $[T \times (K+2)L]$ matrix $\boldsymbol{M}$ of observed data satisfies $\mathrm{rk}(\boldsymbol{M}) = (K+2)L$. However, since in practice $\boldsymbol{M}$ is replaced by sample moment covariances, this issue needs to be determined using a formal statistical test. Camba-Méndez and Kapetanios (2008) analyze a wide range of statistical methods for testing the rank of a matrix.[10] In the Monte Carlo section that follows, we investigate

[10] See also the commonly used procedure of Kleibergen and Paap (2006).

within our set up the properties of the test statistic proposed by Robin and Smith (2000), described in the Appendix, which is relatively simple to implement and has the advantage that it does not require the variance-covariance matrix of $\boldsymbol{M}$ to have full rank, or that its rank is known. The results demonstrate that the method works very well.

3.4. Observed Factors

Consider a model that contains both observed and unobserved factors, that is, in vector form we now have

$$\boldsymbol{y}_i = \alpha \boldsymbol{y}_{i,-1} + \sum_{k=1}^{K} \beta_k \boldsymbol{x}_i^{(k)} + \boldsymbol{F}^o \boldsymbol{\lambda}_i^o + \boldsymbol{F}^u \boldsymbol{\lambda}_i^u + \boldsymbol{\varepsilon}_i, \tag{33}$$

where $\boldsymbol{F}^o$ and $\boldsymbol{F}^u$ denote the observed and unobserved factors, respectively, with dimensions $[T \times L^o]$ and $[T \times L]$. Post-multiplying the model above by $\boldsymbol{w}_i'$, taking expectations and solving for $\boldsymbol{F}^u$ yields

$$\boldsymbol{F}^u = \left(\boldsymbol{M}_y - \alpha \boldsymbol{M}_{y,-1} - \sum_{k=1}^{K} \beta_k \boldsymbol{M}_x^{(k)} - \boldsymbol{F}^o \left(\boldsymbol{G}_w^o \right)' \right) \left((\boldsymbol{G}_w^u)' \right)^{-1},$$

where $\boldsymbol{G}_w^o = \mathrm{E}\left(\boldsymbol{w}_i (\boldsymbol{\lambda}_i^o)' \right)$ is an $[L \times L^o]$ matrix. Plugging this expression into the original model yields

$$\boldsymbol{y}_i = \alpha \boldsymbol{y}_{i,-1} + \sum_{k=1}^{K} \beta_k \boldsymbol{y}_i^{(k)} + \boldsymbol{F}^o \widetilde{\boldsymbol{\lambda}}_i^o + \left(\boldsymbol{M}_y - \alpha \boldsymbol{M}_{y,-1} - \sum_{k=1}^{K} \beta_k \boldsymbol{M}_x^{(k)} \right) \widetilde{\boldsymbol{\lambda}}_i^u + \boldsymbol{\varepsilon}_i,$$

where $\widetilde{\boldsymbol{\lambda}}_i^o = \boldsymbol{\lambda}_i^o - (\boldsymbol{G}_w^o)' \left((\boldsymbol{G}_w^u)' \right)^{-1} \boldsymbol{\lambda}_i^u$ is $[L^o \times 1]$, and $\widetilde{\boldsymbol{\lambda}}_i^u = \left((\boldsymbol{G}_w^u)' \right)^{-1} \boldsymbol{\lambda}_i^u$ is $[L \times 1]$, or

$$\begin{aligned} \boldsymbol{y}_i \;&=\; \alpha \boldsymbol{y}_{i,-1} + \sum_{k=1}^{K} \beta_k \boldsymbol{x}_i^{(k)} + \left(\widetilde{\boldsymbol{\lambda}}_i^o \otimes \boldsymbol{I}_T \right)' \mathrm{vec}\left(\boldsymbol{F}^o \right) \\ &\quad + \left(\widetilde{\boldsymbol{\lambda}}_i^u \otimes \boldsymbol{I}_T \right)' \mathrm{vec}\left(\boldsymbol{M}_y - \alpha \boldsymbol{M}_{y,-1} - \sum_{k=1}^{K} \beta_k \boldsymbol{M}_x^{(k)} \right) + \boldsymbol{\varepsilon}_i. \end{aligned} \tag{34}$$

Thus, multiplying the expression above by $\boldsymbol{Z}_i'$ and taking expectations yields

$$\begin{aligned} \boldsymbol{m} = \alpha \boldsymbol{m}_{-1} &+ \sum_{k=1}^{K} \beta_k \boldsymbol{m}_k + \boldsymbol{S}\, \mathrm{vec}\left(\boldsymbol{F}^o \otimes \boldsymbol{I}_d \right) \boldsymbol{g}^o \\ &+ \boldsymbol{S} \left[\left(\boldsymbol{M}_y - \alpha \boldsymbol{M}_{y,-1} - \sum_{k=1}^{K} \beta_k \left(\boldsymbol{M}_x^{(k)} \right) \right) \otimes \boldsymbol{I}_d \right] \boldsymbol{g}^u, \end{aligned}$$

where $\boldsymbol{G}^o = \mathrm{E}\left(\boldsymbol{z}_i \left(\widetilde{\boldsymbol{\lambda}}_i^o \right)' \right)$, $\boldsymbol{g}^o = \mathrm{vec}\left(\boldsymbol{G}^o \right)$, $\boldsymbol{G}^u = \mathrm{E}\left(\boldsymbol{z}_i \left(\widetilde{\boldsymbol{\lambda}}_i^u \right)' \right)$ and $\boldsymbol{g}^u = \mathrm{vec}\left(\boldsymbol{G}^u \right)$.

One can easily see that the total number of identified parameters is given by

$$\#parameters = K + 1 + d \times (L + L^o) - \xi(L + L^o, \tau(T, 1), \ldots, \tau(T, K)).$$

Analogously to the model without observed factor the estimating equations above can be easily linearized by setting $\boldsymbol{G}_0^u \equiv -\alpha \boldsymbol{G}^u$ and $\boldsymbol{G}_k^u \equiv -\beta_k \boldsymbol{G}^u$. It is clear that the presence of observed factors in the model does not increase the number of parameters following linearization. As a result, the total number of factors is just the sum of linearized observed factors and true observed factors, i.e.

$$\tilde{L} = L(K + 2) + L^o. \tag{35}$$

The total number of non-redundant moment conditions and identifiable parameters follows directly from (32) using the new definition of $\tilde{L}$.

3.5. Selecting the number of factors

So far the true number of factors has been treated as known. However, in practice this quantity is typically unknown and needs to be determined from the data. Within our framework the number of factors can be selected consistently using the Schwartz information criterion, as proposed originally by Ahn et al. (2013). This is formalized in the following proposition:

Proposition 2. *Let $Q_N\left(\widehat{\boldsymbol{\theta}}\left(\boldsymbol{\Omega}_N\right)\right)$ be the value of the objective function evaluated at $\widehat{\boldsymbol{\theta}}$ given $\boldsymbol{\Omega}_N$ and the observed data:*

$$Q_N\left(\widehat{\boldsymbol{\theta}}\left(\boldsymbol{\Omega}_N\right)\right) = \boldsymbol{\mu}_N'\left(\widehat{\boldsymbol{\theta}}\right)\boldsymbol{\Omega}_N\boldsymbol{\mu}_N\left(\widehat{\boldsymbol{\theta}}\right).$$

Consider the following Schwartz Criterion (BIC):

$$S_N\left(L\right) = N \times Q_N\left(\widehat{\boldsymbol{\theta}}\left(\boldsymbol{\Omega}_N\right)|L\right) - \ln\left(N\right) \times h\left(L\right), \tag{36}$$

where $h\left(L\right) = \varrho \times \varkappa\left(L\right) = O\left(1\right)$, a strictly increasing function of L with $0 < \varrho < \infty$ and $\varkappa\left(L\right) = \zeta - \dim\left(\widehat{\boldsymbol{\theta}}\right)$. Under the set of our assumptions, we have

$$\widehat{L} \xrightarrow{p} L_0 \text{ as } N \to \infty.$$

Proof. *This follows directly from Robertson and Sarafidis (2015).*

The above result implies that in principle the empirical researcher may estimate models with $L = 0, 1, \ldots, L_{max}$, and choose $\widehat{L}$ as the value of L that corresponds to the smallest

BIC value. L_{max} itself can be determined based on the overidentifying restrictions test statistic, provided that the latter is constructed using the optimal GMM weighting matrix. In particular, for $L < L_0$ we have $Q_N\left(\widehat{\boldsymbol{\theta}}\left(\boldsymbol{\Omega}_N\right)|L\right) \xrightarrow{p} \infty$ as N grows large. Therefore, fitting a smaller number of factors than the true number should result in rejecting the null hypothesis of the validity of the instruments with probability approaching one as the sample size increases. On the other hand, for $L = L_0$ the overidentifying restrictions test statistic based on the optimal weighting matrix is asymptotically chi-squared distributed with degrees of freedom equal to the difference between moment conditions and estimable parameters. Hence, provided that the level of significance is adjusted downwards as the sample size increases, the probability of rejecting the null hypothesis at $L = L_0$ approaches zero.

4. Monte Carlo study

Our finite sample study considers a dynamic model with $K = 1$, i.e.

$$y_{i,t} = \alpha y_{i,t-1} + \beta x_{i,t} + u_{i,t}; \quad u_{i,t} = \sum_{\ell=1}^{L} \lambda_{\ell,i} f_{\ell,t} + \varepsilon^y_{i,t}, \quad t = 1, \ldots, T.$$

The processes for $x_{i,t}$ and f_t are given, respectively, by

$$x_{i,t} = \delta y_{i,t-1} + \alpha_x x_{i,t-1} + \sum_{\ell=1}^{L} \gamma_{\ell,i} f_{\ell,t} + \varepsilon^x_{i,t};$$
$$f_{\ell,t} = \alpha_f f_{\ell,t-1} + \sqrt{1-\alpha_f^2}\varepsilon^f_{\ell,t}; \quad \varepsilon^f_{\ell,t} \sim \mathcal{N}(0,1), \quad \forall \ell.$$

The factor loadings are generated as $\lambda_{\ell,i} \sim \mathcal{N}(\mu_\lambda, 1)$ and

$$\gamma_{\ell,i} = \mu_\lambda + \rho(\lambda_{\ell,i} - \mu_\lambda) + \sqrt{1-\rho^2}\upsilon^f_{\ell,i}; \quad \upsilon^f_{\ell,i} \sim \mathcal{N}(0,1) \forall \ell,$$

where ρ denotes the correlation coefficient between the factor loadings of the y and x processes. The parameter μ_λ controls the mean of the factor loadings. Furthermore, the idiosyncratic errors are generated as

$$\varepsilon^y_{i,t} \sim \mathcal{N}\left(0,1\right); \quad \varepsilon^x_{i,t} \sim \mathcal{N}\left(0,\sigma_x^2\right), \quad t \geq 0.$$

The signal-to-noise ratio of the model is defined as follows:

$$SNR \equiv \frac{1}{T}\sum_{t=1}^{T} \frac{\text{var}\left(y_{i,t}|\lambda_{\ell,i}, \gamma_{\ell,i}, \{f_{\ell,s}\}_{s=-S}^{t}\right)}{\text{var}\,\varepsilon^y_{i,t}} - 1.$$

In all designs σ_x^2 is set such that $SNR = 5$. This value lies within the range of values considered in the literature, e.g. Bun and Kiviet (2006) specifies $SNR \in \{3; 9\}$. The initial observation for each i is generated as

$$x_{i,0} = \sum_{\ell=1}^{L} \gamma_{\ell,i} f_{\ell,0} + \varepsilon_{i,0}^{x}; \quad f_0 \sim \mathcal{N}(0,1); \quad y_{i,0} = \sum_{\ell=1}^{L} \lambda_{\ell,i} + \varepsilon_{i,0}^{y},$$

which ensures that the parameter g_0 is non-stochastic; see Robertson et al. (2014) for a related discussion. We consider $N = \{200; 800\}$ and $T = \{4; 8\}$. Furthermore, $\alpha = \{0.4; 0.8\}$ and $\beta = 1 - \alpha$, such that the long run parameter always remains equal to 1. The values of the remaining parameters are as follows: $\rho = \{0; 0.6\}$, $\mu_\lambda = \{0; 1\}$, $\delta = \{0; 0.3\}$, $\alpha_x = 0.6$, $\alpha_f = 0.5$ and $L = 1$. The number of replications performed equals $2,000$ for each design and the factors are drawn in each replication.

4.1. Results

First, we consider three non-linear estimators proposed in this paper, based on different choices for w_i. In particular, NC($\cdot$) and Ny($\cdot$) denote the non-linear GMM estimators that make use of $w_i = 1$ and $w_i = y_{i,0}$ respectively, while NY2($\cdot$) is the corresponding estimator making use of $w_i = y_{i,0}^2$. The digit in ($\cdot$) refers to the one step and two step versions of the estimators. Starting values for the one step non-linear estimators are based on two sets of $\mathcal{U}[0,1]$ random variables both for α and for β. For the two step estimators we also include the one step estimates among the starting values.

The results are reported in the Appendix in terms of median bias and root median square error, which is defined as

$$RMSE = \sqrt{\text{med}\left[(\widehat{\alpha}_r - \alpha)^2\right]},$$

where $\widehat{\alpha}_r$ denotes the value of α obtained in the r^{th} replication using a particular estimator. As an additional measure of dispersion we report the radius of the interval centered on the median containing 80% of the observations, divided by *1.28*. This statistic, which we shall refer to as "quasi-standard deviation" (denoted qStd) provides an estimate of the population standard deviation if the distribution were normal, with the advantage that it is more robust to the occurrence of outliers compared to the usual expression for the standard deviation.

Finally, we report empirical rejection frequencies of the t-test, where nominal size is set equal to 5%. For all two step estimators we report results based on corrected standard errors using the correction formula of Windmeijer (2005). Thus, we differ from other studies (e.g. Robertson and Sarafidis (2015) and Ahn et al. (2013)) that do not consider corrected standard errors for non-linear estimators. Furthermore, for the two step GMM estimators we also report the empirical size of the overidentifying restrictions (J) test statistic.

Tables B1 and B4 (B2 and B5) correspond to the non-linear estimators that make use of $w_i = y_{i,0}^2$ ($w_i = 1$) for $\mu_\lambda = 1$ and $\mu_\lambda = 0$, respectively. For the case where $\mu_\lambda = 1$, both estimators perform very well as there exists very little bias and empirical size is close to the nominal level for both α and β. Moreover, the size of the J test is also close to the 5% level in all circumstances. On the other hand, for $\mu_\lambda = 0$ both estimators are biased and size-distorted, as expected. However, the J test statistic appears to have good power in picking this up. In summary, these estimators perform well when they are consistent and the J test statistic has good power to identify cases where they are not.

Tables B3 and B6 present results for Ny($\cdot$), which makes use of $w_i = y_{i,0}^1$, under the same setup. As expected, the performance of the estimator is largely unaffected by the value of μ_λ. Compared to the estimators that use information from μ_λ, the tests based on Ny($\cdot$) tend to be slightly oversized. Furthermore, for $\mu_\lambda = 1$ Ny($\cdot$) is less efficient than the other two estimators (since it makes use of T less moment conditions) and this is well reflected by the larger values of RMSE and qStd.

Tables B7, B8 and B9 present results on the performance of the linearized estimators for the case where $T = 8$ and $\mu_\lambda = 1$. We do not examine $T = 4$ because in this case the linearized estimators are either infeasible, or exactly identified and thereby the overidentifying restrictions test statistic is infeasible. We can observe that the performance of the linearized estimators remains satisfactory, although it is "pound-for-pound" inferior to the performance of the non-linear estimators, which implies that extra simplicity comes with a cost. The results for $\mu_\lambda = 0$ are similar to those for the non-linear estimators, i.e. the linearized estimators that use $w_i = 1$ and $w_i = y_{i,0}^2$ are biased and size-distorted. We do not report these results to save space (available upon request).

Table 1 presents results with respect to testing for the rank of $\boldsymbol{M}$ using the method proposed by Robin and Smith (2000). We focus on the case where $\rho = 0.6$, $\delta = 0.3$ to save space, whereas $\boldsymbol{M}$ is constructed using $w_i = 1$. The results for the remaining parametrizations are very similar and available upon request. Under our set up, $\mu_\lambda = 0$ implies that $\text{rk}(\boldsymbol{M}) = 0$, whereas for $\mu_\lambda = 1$ we have $\text{rk}(\boldsymbol{M}) = 3$. As we observe, under the null hypothesis, $\mu_\lambda = 0$, empirical size is close to its nominal level in all circumstances. Furthermore under the alternative, power is close to unity regardless of the parametrization

Table 1: Test Results for the Rank of M

N	T	α	Size	Power
200	4	.4	.057	.998
200	4	.8	.053	.999
200	8	.4	.051	1.00
200	8	.8	.052	1.00
800	4	.4	.050	1.00
800	4	.8	.050	1.00
800	8	.4	.049	1.00
800	8	.8	.053	1.00

Note: $\delta = 0.3$, $\rho = 0.6$.

selected. Therefore, we may conclude that testing for the rank of $\boldsymbol{M}$, a matrix that can be trivially estimated from the observed data, appears to be an attractive alternative for verifying whether the full rank assumption on $\boldsymbol{G}_w$ holds true or not.

5. Estimation of the Price Elasticity of Urban Water Demand

5.1. Motivation

The need for establishing and maintaining efficient and sustainable urban water management systems is of paramount importance nowadays, especially because of global warming and the ever-increasing urbanization, which are expected to put pressure on natural resources and ecosystems in the near future. According to a recent report published in 2013 by the United Nations, by 2030 there will be over one billion more people living in large urban centres around the world than today.[11]

Urban water networks are characterised by relatively large infrastructure costs compared to operating costs. Thus, as it is common with many other utilities industries and natural monopolies, urban water usage prices are often regulated with a view to recover the costs of production that would occur in a competitive market plus a rate of return on capital. The aforementioned price markup is often inversely related to the price elasticity of water usage demand, a policy rule that is known as Ramsey pricing.

[11]See http://www.un.org/sg/management/pdf/HLP_P2015_Report.pdf, page 18.

Given that water is an essential good for many purposes, consumer demand is price inelastic. However, the actual degree of sensitivity of consumers to changes in prices bears important implications for public policy and decision making for viable water management systems. For example, planning fixed capital investments in alternative water source generation technologies requires such information.[12] Moreover, possible overestimation (say) of the magnitude of the price elasticity of demand can lead to hundreds of millions in revenue losses to a water utility.[13]

The magnitude of price elasticity is largely an empirical issue and it depends, among other things, on the housing composition within a particular urban centre, the average efficiency of water appliances, and so on. As a result, there is a large number of studies in the literature of urban water management that focuses on the estimation of the price elasticity of demand; see e.g. Arbués et al. (2003) and Araral and Wang (2013) for excellent surveys.[14]

Most of existing research employs a static framework. However, as pointed out by Nauges and Thomas (2003b), a dynamic specification is more appropriate since current water use is likely to be influenced by past use, which is due to habit formation in water consumption such as car washing, showering and garden watering, as well as due to the specific stock of durable goods that exists in a house, such as showerheads, washing machines and so on. Nauges and Thomas (2003b) demonstrate that a dynamic model of water usage can be derived from an intertemporal structural optimization problem of price determination, where local communities have a two-fold objective: the maximization of consumers' welfare and cost recovery.

Therefore, in what follows we are going to estimate a dynamic panel data model of water consumption, controlling for local weather conditions and allowing for multiplicative unobserved heterogeneity, which is represented by a factor structure.

[12] See, for example, the recent technical report titled "Assessment of cost recovery through water pricing", which is published by the European Environment Agency, as well as the 2015 IMF Study on managing water challenges and policy instruments. Both are available on the web: see http://www.bdl.hu/_uploads/assessment_of_cost_recovery_through_water˙pricing.pdf and https://www.imf.org/external/pubs/ft/sdn/2015/sdn1511tn.pdf respectively.

[13] See e.g. the Productivity Commission's Inquiry report titled "Australia's Urban Water Sector" which is available at http://www.pc.gov.au/inquiries/completed/urban-water/report/urban-water-volume1.pdf.

[14] Sibly and Tooth (2015) provide an in-depth recent analysis of supply side issues regarding residential water usage.

5.2. Data and Methodology

We examine the effect of prices on urban water usage demand using multi-household level data for New South Wales, Australia. These data have been made publicly available by Sydney Water Corporation as part of the supporting information provided in the study of residential water use pricing that was undertaken by Abrams, Kumaradevan, Sarafidis, and Spaninks (2012). The data are aggregated to some extent in order to maintain privacy. In particular, each cross-sectional unit represents an average of four to six households, which are located nearby and have similar property size.[15]

Sydney Water is the largest water utility in Australia, serving more than 4 million people, while its area of operations covers around $12,700$ km^2. Our sample consists of a balanced panel of $4,500$ multi-household cross-sectional units, each one being observed over a period of 5 years, 2004-2008 inclusive. The original data set is available on a quarterly basis, however the analysis in this section employs year-specific averages in order to avoid any likely contamination with seasonal variation. This is potentially an important issue, as pointed out by Abrams et al. (2012), because to the extent that water demand for outdoor use is more responsive to prices than for indoor use, one expects that price elasticity is higher during the summer compared to the winter.

Our sample contains owner-occupied houses only, which have property size less than 600 m^2. The reasons are two-fold: first of all, households in NSW face distinct price signals, depending on dwelling type (e.g. houses, maisonettes, apartments) and tenancy status (e.g. owner occupied or tenanted). For instance, households of owner-occupied houses face a strong price signal relative to other households because they receive their water bills directly from the water utility; on the other hand, for tenanted houses the landlord may or may not pass on water usage charges to the tenants, depending for example on whether the property is served by an individual water meter or not. Finally, apartments in NSW are most often served by a common water meter in the same building and thereby households do not get charged directly for their water use. This feature implies that the degree of sensitivity to a given change in price across these types of residence is likely to be rather different (heterogeneous) and thereby estimation methods that are applied to pooled data may not provide consistent estimates of the price coefficient.

Secondly, water consumption may be structurally different for houses that occupy very large areas of land. That is, such properties often have their own storage water tank and possibly access to underground water.

[15]See Abrams et al. (2012) for more information regarding the construction of the data set.

For the same reason, our sample contains households that have not participated in a water appliance efficiency program; one can anticipate a smaller price elasticity of demand for households that have already participated in such programs, as they might exhibit a reduced ability to lower their demand based on higher levels of appliance efficiency (demand hardening) and they might also be operating under a "conservation mindset" at first place.

The model we consider to study the price elasticity of water demand is as follows:

$$\ln(cons_{i,t}) = \alpha \ln(cons_{i,t-1}) + \beta_1 price_{i,t} + \beta_2 temp_{i,t} + \beta_3 rain_{i,t} + u_{i,t}, \;\; u_{it} = \boldsymbol{\lambda}_i' \boldsymbol{f}_t + \varepsilon_{i,t}, \quad (37)$$

where $cons_{i,t}$ denotes average daily water consumption for household i at year t, expressed in thousands of litres of water (kL), $price_{i,t}$ is the average real price (in Australian dollars) paid per kilolitre of water used by household i at time t, while $temp_{i,t}$ and $rain_{i,t}$ denote the average amounts of daily rainfall (mm) and temperature (degrees Celsius) during year t. Finally, $u_{i,t}$ is a composite disturbance that contains a factor component. This structure allows for multiplicative unobserved heterogeneity and nests the popular two way error components model as a special case.

By construction the price variable is endogenous because during the period of the analysis a two-tier pricing scheme was in place in NSW such that consumers paid a higher price when their consumption exceeded a certain threshold level.

The values of the weather variables are individual-specific and they have been determined by the physical proximity of each property to a total of thirteen weather stations that exist across Sydney and are operated by the Bureau of Meteorology. The main reason for this is that weather patterns can vary substantially across NSW and, more specifically, in general there are cooler conditions and more rainfall on the coast compared to many areas that are located inland.

The following table presents some descriptive statistics for the variables of the model. The average (median) daily water usage in the sample is roughly .567 (.515) kL, which indicates that water consumption is skewed to the right; this is expected because there is no upper bound in water consumption (loosely speaking). The between standard deviation of daily water usage is larger than the within standard deviation, which implies that there is more variation in water consumption across households than over time, as expected. Interestingly, the same holds true for temperature. On the other hand, the opposite is true for rainfall, i.e. there appears to exist more variation in rainfall over time than across households within the sample. Finally, for the price variable the between standard deviation is about 10 times smaller than the within deviation, i.e. the largest proportion of variation in price is due to the consecutive, year by year, upward changes set by the NSW Independent Pricing and Regulatory Tribunal (IPART). This indicates the importance of having a large cross-sectional dimension in the sample to be able to identify the effect of price.

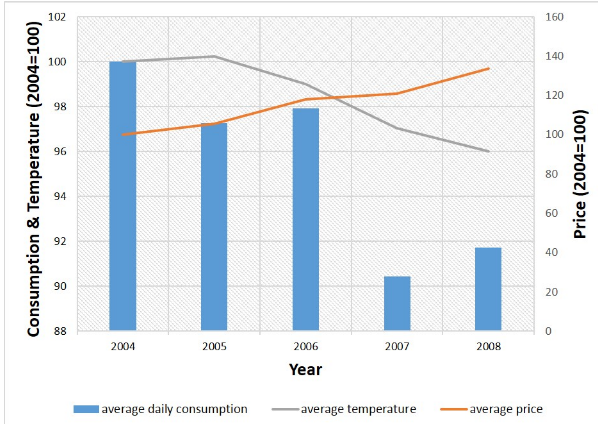
Table 2: Descriptive Statistics

		mean	median	st.dev.	10th perc.	90th perc.
cons.	overall	.567	.515	.328	.203	.977
	between			.303	-.117	.118
	within			.126	.219	.959
rain	overall	2.36	2.16	.739	1.54	3.50
	between			.350	-2.08	2.91
	within			.651	.681	1.05
temp.	overall	23.4	23.7	1.13	21.8	24.5
	between			1.04	21.4	24.2
	within			.437	-.601	.493
price	overall	1.35	1.37	.140	1.17	1.56
	between			.013	1.34	1.36
	within			.139	1.17	1.56

Figure 1 depicts the values of the cross-sectional averages of the variables of the model (except for rain), setting their corresponding 2004 values equal to 100. Therefore, the values of the variables from 2005 onwards are essentially percentage changes relative to the base year. To enhance visualization of the data, the values of water usage and temperature are plotted with respect to the left vertical axis, while those of price are plotted with respect to the right vertical axis. For instance, average water usage in 2007 was roughly 10% lower than 2004. On the other hand, average price in 2008 was roughly 33% higher than 2004. During the period of our analysis, average daily temperature has followed a downward trend overall, whereas prices have steadily gone upwards every single year. At the same time, water usage experienced a significant drop in 2007 and remained much lower in 2008 relative the previous years.

We specify a log-linear functional form for the following two reasons: contrary to

Figure 1: Water Consumption, Price and Temperature



the constant-elasticity (double-log) model, this specification implies that price elasticity depends on the level of price itself, that is, consumers become more sensitive to changes in price the higher the level of price is. This is consistent with utility theory (see, for example Al-Qunaibet and Johnston (1985)). In addition, in comparison to the linear model that has also been popular in the literature, the log-linear specification does not imply the existence of a "choke price" beyond which no water would be demanded from households. This is an important feature of our model because water is an essential product for survival and therefore some water will be consumed even if prices are very high. Notice that our specification also implies that the elasticity of water demand to weather conditions is not constant but depends on the level of temperature and rainfall as well, which is a desirable feature.

We have estimated (37) by fitting models with $L = 0, 1, 2$ factors. In addition, we also estimated (37) (i) by applying first-differences and making use of lagged instruments in levels, which is essentially the popular GMM estimator proposed by Arellano and Bond (1991); and (ii) based on the system GMM estimator proposed e.g. by Arellano and Bover (1995). The number of factors is selected based on the model information criterion in equation (36). Following Ahn et al. (2013) and Robertson and Sarafidis (2015), we set $h(L) = T^{-0.3} \times 0.75 \times df(L)$, where $df(L)$ is the number of degrees of freedom associated with the model fitted with L factors. The performance of this criterion in the context of dynamic panels has been investigated by Robertson and Sarafidis (2015). Starting values for the structural parameters for the non-linear estimators have been obtained using 200 random draws from the standard normal distribution. Optimization is implemented using the iterative procedure described in Sections 2 and 3.

5.3. Results

We estimate the following seven models: $M0$ and MTW_{DIF} (or MTW_{SYS}) denote the models that impose $u_{it} = \varepsilon_{it}$ and $u_{it} = \eta_i + \gamma_t + \varepsilon_{it}$, respectively. That is, $M0$ imposes zero factors, while MTW_{DIF} and MTW_{SYS} impose a two way error components structure. The former is based on the Arellano-Bond estimator and the latter on the System GMM estimator; $M1_c$, $M1_{y_0}$ and $M1_{y_0^2}$ allow for one genuine factor with weights equal to $w_i = 1$, $w_i = y_{i0}$ and $w_i = y_{i0}^2$ respectively; finally, $M2$ allows for two factors with weights given by $\mathbf{w}_i = (1, y_{i0}^2)'$. In all models the price variable is treated as endogenous and is instrumented by appropriate lagged values of the same variable, while the weather variables are treated as exogenous. All models make use of $\zeta = 39$ moment conditions, except for $M1_{y_0}$ that utilises 4 less moment conditions, as well as the Arellano-Bond and System GMM estimators that make use of 17 and 21 moment conditions, respectively.

In Table 3 we summarize the results in terms of the overidentifying restrictions (J) test statistic, its p-value, the number of degrees of freedom for each model, and finally BIC. As it is clear, when we fit either zero factors or a two way error components structure, the p-value of the J statistic is close to zero, which implies that the model is mis-specified. On the other hand, fitting one or two genuine factors leads to failing to reject the null hypothesis that the instruments are valid, which provides evidence that the factor structure is supported by the data over the two way error components structure. Among these specifications $M1_c$ (i.e. $w_i = 1$) corresponds to the smallest BIC value. Therefore in what follows we mainly focus on this particular model.

Table 3: Model Selection

	$M0$	MTW_{DIF}	MTW_{SYS}	$M1_c$	$M1_{y_0}$	$M1_{y_0^2}$	$M2$
J test	156.3	27.6	49.2	21.5	22.9	27.3	3.02
p-value	.000	.002	.000	.369	.291	.128	.933
BIC	10.6	-17.8	-4.91	-61.8	-43.7	-56.6	-30.3
df	35	10	13	20	16	20	8

Table 4 presents the estimation results for the coefficients of the model. The entries in the last row correspond to the long-run price coefficient, computed by dividing the short-run price coefficient, β_1, by one minus the autoregressive parameter, α. The standard error of the long run estimated price coefficient is obtained using the Delta method.[16]

[16] Alternatively, one can use the cross-sectional bootstrap as in Kapetanios (2008), to do inference.

Table 4: Estimation Results for $M1_c$.

	Coef.	**Std. Err.**	**t-ratio**	**p-value**
α	.393	.056	7.06	.000
β_1	-.213	.072	-2.98	.003
β_2	-.006	.009	-0.73	.466
β_3	.040	.010	3.55	.000
$\beta_1/(1-\alpha)$	-.352	.140	-2.50	.012

As we can see, all coefficients have the correct sign. A unit (dollar) increase in the price of water is estimated to cause a reduction in (logged) water consumption of approximately .213 and .352 units in the short- and long-run respectively. Similarly, a unit increase in temperature (rain) is expected to increase (reduce) water consumption by approximately .039 (.006) units in the short-run and .064 (.001) units in the long-run. The coefficient of rainfall is not significantly different from zero, which is consistent with findings in other studies in the literature (see e.g. Abrams et al. (2012)). The value of the autoregressive coefficient is less than .4 and implies that it takes about 2.5 time periods (years) for 90% of the total (i.e. long-run) price effect to be realized, all other things remaining constant.

It is worth noting here that the estimated short-run price coefficient obtained from MTW_{DIF} and MTW_{SYS} is positive and statistically insignificant in both cases. This demonstrates the importance of allowing for a genuine factor component in the model.

We remark that the result of the overidentifying restrictions test statistic for $M1_c$ implies that $\mu_\lambda \neq 0$. As a way of cross-checking this outcome we wish to test for the rank of $\boldsymbol{M}$, computed based on $w_i = 1$, using the procedure described in the Appendix. Given that the data do not have mean zero, an appropriate null hypothesis is $\text{rk}\,(\boldsymbol{M}) = K + 2$. However, since $K+2 > T$ the test is not feasible as it stands. Therefore, to make progress we have implemented the test by constructing $\boldsymbol{M}$ without $rain_{i,t}$. This strategy makes sense because the slope coefficient of this variable is close to zero. The resulting CRT test statistic equals 2.11 and has a p-value close to zero. Hence, we conclude that $\mu_\lambda \neq 0$.

The price elasticity of demand is computed by multiplying the relevant price coefficients with a range of values for price. In Table 5 we present some estimates of the price elasticity at four different values of price; namely, mean and median price, as well as the 10th and 90th percentiles. For example, at the median price within the sample of $1.151 per kL the price elasticity of demand is about .293 per cent in the short-run and .482 per cent in the long-run.

Table 5: Point-wise predicted elasticities for $M1_c$.

	10th perc.	**mean**	**median**	**90th perc.**
price	1.17	1.35	1.37	1.56
SR elasticity	-.249	-.288	-.293	-.333
LR elasticity	-.411	-.475	-.482	-.548

As expected, urban water demand appears to be rather more elastic in the long-run than in the short run, which may be attributed to habit formation and technological constraints of water appliances. This outcome shows that it is important to estimate a dynamic model of urban water demand. In comparison to other studies in the literature, the estimated price elasticity of demand obtained in the present paper is statistically similar to the value obtained by Nauges and Thomas (2003b) (see Table III in their paper) although theirs is derived from the constant-elasticity model using municipal-level data and includes average income but not weather conditions. As a robustness check we have also estimated a model that includes NSW-wide disposable income as a common *observed* factor, with household-specific loadings. These loadings could be interpreted as if they reflected the scalar of proportionality for household i's income over the NSW-wide average income. Thus, for (say) $\gamma_i = 1.2$ household i's income is 1.2 times greater than the average. These results are very similar to those described above and thereby we do not report them.

6. Concluding Remarks

This paper put forward a new methodology that simplifies estimation of dynamic panel data models with multi-factor residuals, for fixed number of time-series observations. The underlying idea is to replace the unobserved factors with (weighted) averages of observed data. This leads to a more parsimonious parametrization of the model. As a result, starting values of the unknown parameters are easy to obtain. The simulation exercise shows that the proposed estimators perform more than satisfactorily and their finite samples properties are well understood.

We hope that the proposed methodology will enhance the application of estimators that allow for multi-factor residuals in panels involving micro level data, and encourage empirical researches to implement these estimators in practice.

References

Abadir, K. M. and J. R. Magnus (2002): "Notation in Econometrics: A Proposal for a Standard," *Econometrics Journal*, 5, 76–90.

Abrams, B., S. Kumaradevan, V. Sarafidis, and F. Spaninks (2012): "An Econometric Assessment to Pricing Sydney's Residential Water Use," *Economic Record*, 88, 89–105.

Ahn, S. C., Y. H. Lee, and P. Schmidt (2001): "GMM Estimation of Linear Panel Data Models with Time-varying Individual Effects," *Journal of Econometrics*, 101, 219–255.

——— (2013): "Panel Data Models with Multiple Time-varying Individual Effects," *Journal of Econometrics*, 174, 1–14.

Al-Qunaibet, M. and R. Johnston (1985): "Municipal Demand for Water in Kuwait: Methodological Issues and Empirical Results," *Water Resources Research*, 21, 433–438.

Araral, E. and Y. Wang (2013): "Water Demand Management: Review of Literature and Comparison in South-East Asia," *International Journal of Water Resources Development*, 29, 434–450.

Arbués, F., M. García-Valinas, and R. Martínez-Espineira (2003): "Estimation of Residential Water Demand: A State-of-the-Art Review," *Journal of Socio-Economics*, 32, 81–102.

Arellano, M. (2003): *Panel Data Econometrics*, Advanced Texts in Econometrics, Oxford University Press.

Arellano, M. and S. Bond (1991): "Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations," *Review of Economic Studies*, 58, 277–297.

Arellano, M. and O. Bover (1995): "Another Look at the Instrumental Variable Estimation of Error-components Models," *Journal of Econometrics*, 68, 29–51.

Bai, J. (2013): "Likelihood Approach to Dynamic Panel Models with Interactive Effects," Working Paper.

Bun, M. J. G. and J. F. Kiviet (2006): "The Effects of Dynamic Feedbacks on LS and MM Estimator Accuracy in Panel Data Models," *Journal of Econometrics*, 132, 409–444.

Bun, M. J. G. and V. Sarafidis (2015): "Dynamic Panel Data Models," in *The Oxford Handbook of Panel Data*, ed. by B. H. Baltagi, Oxford University Press, chap. 3.

Camba-Méndez, G. and G. Kapetanios (2008): "Statistical tests and estimators of the rank of a matrix and their applications in econometric modelling," ECB working paper No 850.

Chamberlain, G. (1982): "Multivariate Regression Models for Panel Data," *Journal of Econometrics*, 18, 5–46.

Hayakawa, K. (2012): "GMM Estimation of Short Dynamic Panel Data Model with Interactive Fixed Effects," *Journal of the Japan Statistical Society*, 42, 109–123.

Hayakawa, K., H. M. Pesaran, and L. V. Smith (2014): "Transformed Maximum Likelihood Estimation of Short Dynamic Panel Data Models with Interactive Effectects," Working Paper.

Holtz-Eakin, D., W. K. Newey, and H. S. Rosen (1988): "Estimating Vector Autoregressions with Panel Data," *Econometrica*, 56, 1371–1395.

Juodis, A. and V. Sarafidis (2014): "Fixed T Dynamic Panel Data Estimators with Multi-Factor Errors," UvA-Econometrics working paper 2014/07.

Kapetanios, G. (2008): "A Bootstrap Procedure for Panel Data Sets with Many Cross-sectional Units," *Econometrics Journal*, 11, 377–395.

Karabiyik, H., J. P. Urbain, and J. Westerlund (2014): "CCE Estimation of Factor-Augmented Regression Models with More Factors than Observables," GSBE Research Memorandum RM/14/07.

Kleibergen, F. R. and R. Paap (2006): "Generalized Reduced Rank Tests Using the Singular Value Decomposition," *Journal of Econometrics*, 133, 97–126.

Kruiniger, H. (2008): "Not So Fixed Effects: Correlated Structural Breaks in Panel Data," Unpublished manuscript, Queen Marry, University of London.

Nauges, C. and A. Thomas (2003a): "Consistent Estimation of Dynamic Panel Data Models with Time-varying Individual Effects," *Annales d'Economie et de Statistique*, 70, 54–75.

——— (2003b): "Long-Run Study of Residential Water Consumption," *Environmental and Resource Economics*, 26, 25–43.

Pesaran, H. M. (2006): "Estimation and Inference in Large Heterogeneous Panels with a Multifactor Error Structure," *Econometrica*, 74, 967–1012.

Robertson, D. and V. Sarafidis (2015): "IV Estimation of Panels with Factor Residuals," *Journal of Econometrics*, 185, 526–541.

Robertson, D., V. Sarafidis, and J. Westerlund (2014): "GMM Unit Root Inference in Generally Trending and Cross-Correlated Dynamic Panels," Working Paper.

Robin, J. M. and R. J. Smith (2000): "Tests of Rank," *Econometric Theory*, 16, 151–175.

Sibly, H. and R. Tooth (2015): "Managing Water Variability Issues," in *Understanding and Managing Urban Water in Transition*, ed. by Q. Grafton, K. Daniel, C. Nauges, J. Rinaudo, and N. Chan, Springer, chap. 18, 383–400.

Windmeijer, F. (2005): "A Finite Sample Correction for the Variance of Linear Efficient Two-Step GMM Estimators," *Journal of Econometrics*, 126, 25–51.

Appendix A. Theoretical Results

Appendix A.1. Derivation of the S.E. correction for the ARX(1) model

For convenience of derivations we define $\beta_0 \equiv \alpha$ and $\boldsymbol{X}_0 \equiv \boldsymbol{Y}_{-1}$. One can similarly accommodate further lags of $y_{i,t}$, but for simplicity we stick to the ARX(1) model. Next, we rewrite the moment conditions as

$$\boldsymbol{\mu}_N(\boldsymbol{\theta}) = \frac{1}{N}\boldsymbol{S}\,\mathrm{vec}\left((\boldsymbol{\Xi} - \boldsymbol{W}\boldsymbol{G}')'\left(\boldsymbol{Y} - \sum_{k=0}^{K}\beta_k\boldsymbol{X}_k\right)\right). \tag{A.1}$$

Here all $\boldsymbol{Y}, \boldsymbol{Y}_{-1}, \boldsymbol{X}_1, \ldots \boldsymbol{X}_K$ are $[N \times T]$, $\boldsymbol{\Xi}$ is $[N \times d]$ and $\boldsymbol{W}$ is $[N \times L]$, with typical row elements $\boldsymbol{y}_i'$, $\boldsymbol{y}_{i,-1}'$, $\boldsymbol{x}_i^{(k)\prime}$, $\boldsymbol{z}_i'$ and $\boldsymbol{w}_i'$ respectively. $\boldsymbol{G}$ is of dimension $[d \times L]$, while recall that $\boldsymbol{S}$ is of dimension $[\zeta \times Td]$. The $[\zeta \times (K+1+dL)]$ Jacobian matrix $\widehat{\boldsymbol{\Gamma}}_N(\boldsymbol{\theta})$ of these moment conditions is given by:

$$\widehat{\boldsymbol{\Gamma}}_N(\boldsymbol{\theta}) = \left(\ \widehat{\boldsymbol{\Gamma}}_N^{(1)}(\boldsymbol{\theta}), \quad \widehat{\boldsymbol{\Gamma}}_N^{(2)}(\boldsymbol{\theta}) \ \right).$$

Here

$$\widehat{\boldsymbol{\Gamma}}_N^{(1)}(\boldsymbol{\theta}) = -\frac{1}{N}\boldsymbol{S}\left(\ \mathrm{vec}\left((\boldsymbol{\Xi} - \boldsymbol{W}\boldsymbol{G}')'\boldsymbol{X}_0\right), \quad \ldots \quad , \mathrm{vec}\left((\boldsymbol{\Xi} - \boldsymbol{W}\boldsymbol{G}')'\boldsymbol{X}_K\right) \ \right);$$
$$\widehat{\boldsymbol{\Gamma}}_N^{(2)}(\boldsymbol{\theta}) = -\frac{1}{N}\boldsymbol{S}\left(\left(\left(\boldsymbol{Y} - \sum_{k=0}^{K}\beta_k\boldsymbol{X}_k\right)'\boldsymbol{W}\right) \otimes \boldsymbol{I}_d\right).$$

Finally, we can derive the second-derivative ($\boldsymbol{G}_N$, in the notation of Windmeijer (2005)) matrix which is defined in the following way:

$$\mathrm{dvec}(\widehat{\boldsymbol{\Gamma}}_N(\boldsymbol{\theta})) = \boldsymbol{H}_N \, \mathrm{d}\boldsymbol{\theta},$$

where $\boldsymbol{H}_N$ is $[\zeta(K+1+dL) \times (K+1+dL)]$ It is straightforward to derive the above term given that both blocks of the $\widehat{\boldsymbol{\Gamma}}_N(\boldsymbol{\theta})$ matrix are linear in parameters. As a result we have

$$\boldsymbol{H}_N = \begin{pmatrix} \mathbf{O} & \boldsymbol{H}_N^{(1,2)} \\ \boldsymbol{H}_N^{(2,1)} & \mathbf{O} \end{pmatrix},$$

where

$$\boldsymbol{H}_N^{(1,2)} = \frac{1}{N}(\boldsymbol{S} \otimes \boldsymbol{I}_{K+1}) \begin{pmatrix} (\boldsymbol{X}_0'\boldsymbol{W}) \otimes \boldsymbol{I}_d \\ \vdots \\ (\boldsymbol{X}_K'\boldsymbol{W}) \otimes \boldsymbol{I}_d \end{pmatrix},$$
$$\boldsymbol{H}_N^{(2,1)} = \frac{1}{N}\left(\ \mathrm{vec}\left(\boldsymbol{S}\left((\boldsymbol{X}_0'\boldsymbol{W}) \otimes \boldsymbol{I}_d\right)\right), \quad \ldots, \quad \mathrm{vec}\left(\boldsymbol{S}\left((\boldsymbol{X}_K'\boldsymbol{W}) \otimes \boldsymbol{I}_d\right)\right) \ \right),$$

are $[\zeta(K+1) \times dL]$ and $[\zeta(dL) \times (K+1)]$ respectively. The $\widehat{\boldsymbol{\Gamma}}_N(\boldsymbol{\theta})$ and $\boldsymbol{H}_N$ matrices can be plugged into the formula (2.5) of Windmeijer (2005) (assuming one uses $\boldsymbol{I}_\zeta$ as an identity matrix in the first step):

$$\begin{aligned}
\widehat{\mathrm{var}}_c(\hat{\boldsymbol{\theta}}_2) = {} & \frac{1}{N}\boldsymbol{A}^{-1}_{\hat{\boldsymbol{\theta}}_2,\boldsymbol{\Omega}_N(\hat{\boldsymbol{\theta}}_1)}\widehat{\boldsymbol{\Gamma}}_N(\hat{\boldsymbol{\theta}}_2)'\boldsymbol{\Omega}_N^{-1}(\hat{\boldsymbol{\theta}}_1)\widehat{\boldsymbol{\Gamma}}_N(\hat{\boldsymbol{\theta}}_2)\boldsymbol{A}^{-1}_{\hat{\boldsymbol{\theta}}_2,\boldsymbol{\Omega}_N(\hat{\boldsymbol{\theta}}_1)} \\
& + \frac{1}{N}\boldsymbol{D}_{\hat{\boldsymbol{\theta}}_2,\boldsymbol{\Omega}_N(\hat{\boldsymbol{\theta}}_1)}\boldsymbol{A}^{-1}_{\hat{\boldsymbol{\theta}}_1}\widehat{\boldsymbol{\Gamma}}_N(\hat{\boldsymbol{\theta}}_1)'\widehat{\boldsymbol{\Gamma}}_N(\hat{\boldsymbol{\theta}}_2)\boldsymbol{A}^{-1}_{\hat{\boldsymbol{\theta}}_2,\boldsymbol{\Omega}_N(\hat{\boldsymbol{\theta}}_1)} \\
& + \frac{1}{N}\left(\boldsymbol{D}_{\hat{\boldsymbol{\theta}}_2,\boldsymbol{\Omega}_N(\hat{\boldsymbol{\theta}}_1)}\boldsymbol{A}^{-1}_{\hat{\boldsymbol{\theta}}_1}\widehat{\boldsymbol{\Gamma}}_N(\hat{\boldsymbol{\theta}}_1)'\widehat{\boldsymbol{\Gamma}}_N(\hat{\boldsymbol{\theta}}_2)\boldsymbol{A}^{-1}_{\hat{\boldsymbol{\theta}}_2,\boldsymbol{\Omega}_N(\hat{\boldsymbol{\theta}}_1)}\right)' \\
& + \boldsymbol{D}_{\hat{\boldsymbol{\theta}}_2,\boldsymbol{\Omega}_N(\hat{\boldsymbol{\theta}}_1)}\widehat{\mathrm{var}}(\hat{\boldsymbol{\theta}}_1)\boldsymbol{D}'_{\hat{\boldsymbol{\theta}}_2,\boldsymbol{\Omega}_N(\hat{\boldsymbol{\theta}}_1)},
\end{aligned}$$

where $\hat{\boldsymbol{\theta}}_2$ is the two-step GMM estimator with optimal weighting matrix $\boldsymbol{\Omega}_N(\hat{\boldsymbol{\theta}}_1)$ based on the residuals from the firs-step. While,

$$\begin{aligned}
\boldsymbol{A}_{\hat{\boldsymbol{\theta}}_2,\boldsymbol{\Omega}_N(\hat{\boldsymbol{\theta}}_1)} &= \widehat{\boldsymbol{\Gamma}}_N(\hat{\boldsymbol{\theta}}_2)'\boldsymbol{\Omega}_N^{-1}(\hat{\boldsymbol{\theta}}_1)\widehat{\boldsymbol{\Gamma}}_N(\hat{\boldsymbol{\theta}}_2) + \boldsymbol{H}_N'\left(\boldsymbol{I}_{K+1+dL}\otimes\boldsymbol{\Omega}_N^{-1}(\hat{\boldsymbol{\theta}}_1)\boldsymbol{\mu}_N(\hat{\boldsymbol{\theta}}_2)\right);\\
\boldsymbol{A}_{\hat{\boldsymbol{\theta}}_1} &= \widehat{\boldsymbol{\Gamma}}_N(\hat{\boldsymbol{\theta}}_1)'\widehat{\boldsymbol{\Gamma}}_N(\hat{\boldsymbol{\theta}}_1);\\
\widehat{\mathrm{var}}(\hat{\boldsymbol{\theta}}_1) &= \frac{1}{N}\boldsymbol{A}_{\hat{\boldsymbol{\theta}}_1}^{-1}\widehat{\boldsymbol{\Gamma}}_N(\hat{\boldsymbol{\theta}}_1)'\boldsymbol{\Omega}_N^{-1}(\hat{\boldsymbol{\theta}}_1)\widehat{\boldsymbol{\Gamma}}_N(\hat{\boldsymbol{\theta}}_1)\boldsymbol{A}_{\hat{\boldsymbol{\theta}}_1}^{-1}.
\end{aligned}$$

Note that we slightly deviate from the formula used in Windmeijer (2005) to calculate the $\widehat{\mathrm{var}}(\hat{\boldsymbol{\theta}}_1)$ because we ignored the second term to simplify computations. Alternatively, one can consider $\boldsymbol{A}_{\hat{\boldsymbol{\theta}}_1} = \widehat{\boldsymbol{\Gamma}}_N(\hat{\boldsymbol{\theta}}_1)'\widehat{\boldsymbol{\Gamma}}_N(\hat{\boldsymbol{\theta}}_1) + \boldsymbol{H}_N'\left(\boldsymbol{I}_{K+1+dL}\otimes\boldsymbol{\mu}_N(\hat{\boldsymbol{\theta}}_1)\right)$. The j^{th} column of the $[(K+1+dL)\times(K+1+dL)]$ matrix $\boldsymbol{D}_{\boldsymbol{\theta},\boldsymbol{\Omega}_N(\boldsymbol{\theta})}$, say $\boldsymbol{D}_{\hat{\boldsymbol{\theta}}_2,\boldsymbol{\Omega}_N(\hat{\boldsymbol{\theta}}_1)}[,j]$ is given by the following expression:

$$\begin{aligned}
\boldsymbol{D}_{\hat{\boldsymbol{\theta}}_2,\boldsymbol{\Omega}_N(\hat{\boldsymbol{\theta}}_1)}[,j] &= \boldsymbol{A}^{-1}_{\hat{\boldsymbol{\theta}}_2,\boldsymbol{\Omega}_N(\hat{\boldsymbol{\theta}}_1)}\widehat{\boldsymbol{\Gamma}}_N(\hat{\boldsymbol{\theta}}_2)'\boldsymbol{\Omega}_N^{-1}(\hat{\boldsymbol{\theta}}_1)\left.\frac{\partial\boldsymbol{\Omega}_N(\boldsymbol{\theta})}{\partial\theta_j}\right|_{\hat{\boldsymbol{\theta}}_1}\boldsymbol{\Omega}_N^{-1}(\hat{\boldsymbol{\theta}}_1)\boldsymbol{\mu}_N(\hat{\boldsymbol{\theta}}_2);\\
\frac{\partial\boldsymbol{\Omega}_N(\boldsymbol{\theta})}{\partial\theta_j} &= \frac{1}{N}\sum_{i=1}^{N}\left(\frac{\partial\boldsymbol{\mu}_i(\boldsymbol{\theta})}{\partial\theta_j}\boldsymbol{\mu}_i(\boldsymbol{\theta})' + \boldsymbol{\mu}_i(\boldsymbol{\theta})\frac{\partial\boldsymbol{\mu}_i(\boldsymbol{\theta})'}{\partial\theta_j}\right).
\end{aligned}$$

Results for the model with observed factors follows analogously by noting that in that case one has

$$\boldsymbol{\mu}_N(\boldsymbol{\theta}) = \boldsymbol{S}\,\mathrm{vec}\left(\frac{1}{N}\left(\boldsymbol{\Xi}-\boldsymbol{W}\boldsymbol{G}'\right)'\left(\boldsymbol{Y}-\sum_{k=0}^{K}\beta_k\boldsymbol{X}_k\right)-\boldsymbol{G}^o(\boldsymbol{F}^o)'\right). \tag{A.2}$$

Appendix A.2. Proof of Theorem 1

Proof. For notational simplicity we consider the case where $\alpha = \beta_2 = \ldots \beta_K = 0$. Therefore, with a slight abuse of notation we can write

$$\boldsymbol{M}_y = \beta\boldsymbol{M}_x + \boldsymbol{F}\boldsymbol{G}_w = \beta\boldsymbol{M}_x + \boldsymbol{V},$$

where $\boldsymbol{F}\boldsymbol{G}_w = \boldsymbol{V}$. Let $\boldsymbol{M} \equiv (\boldsymbol{M}_y, \boldsymbol{M}_x)$, a $[T\times 2L]$ matrix. Therefore, solving for $\boldsymbol{V}$ yields

$$\boldsymbol{V} = (\boldsymbol{M}_y,\boldsymbol{M}_x)(\boldsymbol{I}_L, -\beta\boldsymbol{I}_L) = \boldsymbol{M}\boldsymbol{B}.$$

It is clear that $\mathrm{rank}(\boldsymbol{B}) = L$.

Suppose that matrix $\boldsymbol{M}$ has full rank, i.e. the column space of $\boldsymbol{M}$ spans $\boldsymbol{R}^{2L}$. This means that there exists a submatrix of $\boldsymbol{M}$, given by $\boldsymbol{M}_s = \boldsymbol{S}\boldsymbol{M}$, where $\boldsymbol{S}$ is a $[2L\times T]$ selector matrix, such that $\boldsymbol{M}_s$ is non-singular. Hence, using a standard result of rank equality for non-singular matrices (see e.g. page 13 in Horn and Johnson, 1995) we have

$$\mathrm{rank}\left(\boldsymbol{M}_s\boldsymbol{B}\right) = \mathrm{rank}\left(\boldsymbol{B}\right) = L.$$

However, due to the fact that

$$\boldsymbol{M}_s\boldsymbol{B} = \boldsymbol{S}\boldsymbol{M}\boldsymbol{B} = \boldsymbol{S}\boldsymbol{V},$$

we also have

$$\operatorname{rank}(\boldsymbol{S}\boldsymbol{V}) = L.$$

Therefore, since

$$\operatorname{rank}(\boldsymbol{S}\boldsymbol{V}) \leq \operatorname{rank}(\boldsymbol{V}) = \operatorname{rank}(\boldsymbol{F}\boldsymbol{G}_w) \leq \operatorname{rank}(\boldsymbol{G}_w),$$

it follows that $\text{rank}(\boldsymbol{G}_w) = L$.

For the general case, one can write

$$\boldsymbol{M}_y = \alpha\boldsymbol{M}_{y,-1} + \sum_{k=1}^{K}\beta_k\boldsymbol{M}_x^{(k)} + \boldsymbol{V},$$

and solve in terms of $\boldsymbol{V} = \boldsymbol{M}\boldsymbol{B}$, where $\boldsymbol{M} \equiv \left(\boldsymbol{M}_y, \boldsymbol{M}_{y,-1}, \boldsymbol{M}_x^{(1)}, \ldots, \boldsymbol{M}_x^{(K)}\right)$. The proof then follows in an identical way. *QED* □

Appendix A.3. Testing for the rank of $\boldsymbol{M}$

The method proposed by Robin and Smith (2000) focuses on the eigenvalues of quadratic forms of a matrix. In particular, let $\widehat{\boldsymbol{M}}$ denote the sample counterpart of $\boldsymbol{M}$ and $\boldsymbol{A} \equiv \widehat{\boldsymbol{M}}\widehat{\boldsymbol{M}}'$. Consider the following statistic:

$$CRT = N\sum_{\ell=\varrho+1}^{(K+2)L}\varpi_\ell, \tag{A.3}$$

where ϖ_ℓ are the eigenvalues of the $[T \times T]$ matrix $\boldsymbol{A}$ in descending order and ϱ denotes the rank of $\boldsymbol{M}$. Under the null hypothesis, the above statistic converges in distribution to a weighted sum of independent χ_1^2 random variables. The weights are given by the largest eigenvalues – denoted by τ_j with $j = 1, \ldots, (T-\varrho)((K+2)L-\varrho)$ – of $\left(\boldsymbol{D}_\varrho' \otimes \boldsymbol{C}_\varrho'\right)\boldsymbol{V}(\boldsymbol{D}_\varrho \otimes \boldsymbol{C}_\varrho)$, where $\boldsymbol{D}_\varrho$ and $\boldsymbol{C}_\varrho$ denote the eigenvectors corresponding to the $(K+2)L-\varrho$ and $T-\varrho$ smallest eigenvalues of $\boldsymbol{M}'\boldsymbol{M}$ and $\boldsymbol{A}$, respectively, while $\boldsymbol{V}$ is such that

$$\sqrt{N}\operatorname{vec}\left(\widehat{\boldsymbol{M}} - \boldsymbol{M}\right) \xrightarrow{d} N(\boldsymbol{0}, \boldsymbol{V}). \tag{A.4}$$

Appendix B. Monte Carlo results

Table B1: Nonlinear GMM estimator using $w_i = y_{i,0}^2$ and $\mu_\lambda = 1$

Designs					GMM 1 step								GMM 2 step								
					α				β				α				β				J
N	T	α	ρ	δ	Bias	RMSE	qStd	Size	Bias	RMSE	qStd	Size	Bias	RMSE	qStd	Size	Bias	RMSE	qStd	Size	Size
200	4	.4	.0	.0	-.002	.022	.070	.062	.001	.016	.051	.064	.000	.020	.060	.057	.001	.015	.044	.065	.046
200	4	.4	.0	.3	-.005	.036	.113	.072	.003	.026	.079	.069	-.001	.027	.081	.061	.002	.020	.059	.070	.044
200	4	.4	.6	.0	-.002	.024	.073	.062	.002	.017	.051	.060	-.001	.021	.061	.059	.001	.014	.044	.062	.044
200	4	.4	.6	.3	-.005	.039	.121	.075	.005	.026	.080	.061	-.003	.028	.084	.062	.002	.019	.059	.065	.043
200	4	.8	.0	.0	-.003	.027	.079	.062	.000	.009	.026	.063	-.001	.022	.067	.062	.001	.008	.023	.064	.043
200	4	.8	.0	.3	-.005	.033	.103	.081	.001	.011	.033	.070	-.003	.025	.076	.061	.001	.009	.027	.068	.045
200	4	.8	.6	.0	-.003	.028	.083	.064	.001	.009	.026	.069	-.002	.023	.068	.064	.001	.007	.023	.064	.042
200	4	.8	.6	.3	-.005	.035	.106	.079	.001	.011	.032	.075	-.003	.026	.079	.062	.001	.009	.027	.067	.045
200	8	.4	.0	.0	-.002	.014	.043	.051	.003	.014	.041	.056	-.001	.011	.034	.089	.001	.011	.033	.116	.034
200	8	.4	.0	.3	-.014	.036	.107	.071	.013	.034	.102	.077	-.005	.020	.061	.132	.005	.019	.057	.149	.033
200	8	.4	.6	.0	-.003	.015	.046	.049	.003	.013	.040	.061	-.001	.012	.036	.096	.001	.011	.033	.120	.032
200	8	.4	.6	.3	-.017	.037	.113	.074	.015	.034	.106	.078	-.006	.021	.062	.133	.006	.019	.058	.150	.030
200	8	.8	.0	.0	-.002	.013	.040	.050	.001	.007	.021	.056	-.001	.011	.035	.095	.001	.006	.019	.114	.034
200	8	.8	.0	.3	-.004	.021	.063	.053	.003	.012	.038	.066	-.002	.014	.045	.107	.001	.010	.029	.125	.034
200	8	.8	.6	.0	-.002	.014	.042	.052	.001	.007	.021	.057	.000	.012	.035	.094	.001	.006	.019	.114	.030
200	8	.8	.6	.3	-.006	.022	.066	.059	.004	.013	.039	.064	-.003	.015	.046	.107	.002	.010	.029	.126	.032
800	4	.4	.0	.0	-.001	.012	.034	.059	.001	.008	.025	.054	.000	.010	.029	.059	.000	.007	.022	.062	.040
800	4	.4	.0	.3	-.002	.018	.055	.051	.001	.012	.038	.055	.000	.013	.039	.053	.000	.009	.028	.056	.042
800	4	.4	.6	.0	.000	.012	.036	.056	.001	.008	.025	.054	.000	.010	.030	.059	.000	.007	.021	.057	.043
800	4	.4	.6	.3	-.002	.018	.058	.055	.002	.013	.039	.055	.000	.013	.040	.056	.000	.009	.028	.058	.042
800	4	.8	.0	.0	-.001	.014	.041	.055	.000	.004	.013	.054	.000	.010	.031	.054	.000	.004	.011	.051	.042
800	4	.8	.0	.3	-.001	.017	.051	.056	.000	.005	.016	.061	-.001	.012	.036	.048	.000	.004	.013	.046	.040
800	4	.8	.6	.0	-.001	.015	.042	.054	.000	.004	.013	.052	.000	.010	.032	.049	.000	.004	.011	.053	.042
800	4	.8	.6	.3	-.001	.018	.054	.055	.001	.005	.017	.052	-.001	.012	.036	.046	.000	.004	.012	.048	.041
800	8	.4	.0	.0	-.001	.007	.021	.049	.001	.006	.019	.060	.000	.005	.015	.059	.000	.005	.015	.073	.054
800	8	.4	.0	.3	-.004	.017	.054	.060	.004	.016	.051	.063	-.001	.008	.024	.072	.001	.008	.023	.077	.050
800	8	.4	.6	.0	-.001	.007	.022	.044	.001	.006	.019	.059	.000	.005	.016	.059	.000	.005	.014	.072	.053
800	8	.4	.6	.3	-.004	.017	.056	.059	.004	.016	.051	.054	-.001	.008	.024	.071	.001	.008	.023	.074	.050
800	8	.8	.0	.0	-.001	.007	.020	.045	.000	.003	.010	.064	.000	.005	.016	.056	.000	.003	.009	.076	.051
800	8	.8	.0	.3	-.001	.010	.033	.048	.001	.006	.019	.058	.000	.006	.019	.057	.000	.004	.013	.077	.051
800	8	.8	.6	.0	-.001	.007	.021	.045	.000	.003	.010	.063	.000	.006	.016	.055	.000	.003	.008	.073	.053
800	8	.8	.6	.3	-.001	.011	.035	.049	.001	.006	.020	.057	.000	.007	.020	.058	.000	.004	.012	.075	.049

Table B2: Nonlinear GMM estimator using $w_i = y_{i,0}^0$ and $\mu_\lambda = 1$

Designs					GMM 1 step								GMM 2 step								
					α				β				α				β				J
N	T	α	ρ	δ	Bias	RMSE	qStd	Size	Bias	RMSE	qStd	Size	Bias	RMSE	qStd	Size	Bias	RMSE	qStd	Size	Size
200	4	.4	.0	.0	-.001	.022	.067	.060	.000	.016	.047	.061	.001	.019	.058	.057	.000	.014	.043	.067	.051
200	4	.4	.0	.3	-.002	.034	.101	.057	.001	.023	.070	.060	.000	.026	.077	.066	.001	.019	.056	.066	.051
200	4	.4	.6	.0	-.003	.024	.072	.066	.002	.017	.049	.056	.000	.020	.058	.056	.001	.014	.043	.061	.051
200	4	.4	.6	.3	-.008	.038	.113	.061	.004	.026	.076	.062	-.001	.027	.080	.061	.002	.018	.056	.066	.049
200	4	.8	.0	.0	-.002	.026	.080	.065	.000	.008	.025	.066	.000	.021	.062	.065	.000	.007	.022	.063	.047
200	4	.8	.0	.3	-.003	.033	.094	.068	.000	.010	.032	.066	-.002	.023	.069	.064	.001	.009	.026	.066	.048
200	4	.8	.6	.0	-.003	.026	.082	.066	.001	.008	.025	.068	-.001	.021	.063	.070	.001	.007	.022	.062	.045
200	4	.8	.6	.3	-.004	.032	.099	.067	.001	.011	.032	.064	-.002	.024	.071	.068	.001	.008	.026	.065	.048
200	8	.4	.0	.0	-.001	.013	.039	.058	.000	.012	.035	.058	.000	.011	.034	.097	.000	.010	.031	.116	.035
200	8	.4	.0	.3	-.005	.028	.085	.060	.004	.025	.077	.061	-.001	.018	.055	.118	.001	.016	.051	.129	.032
200	8	.4	.6	.0	-.002	.015	.045	.058	.002	.012	.038	.061	.000	.012	.036	.103	.001	.010	.032	.117	.035
200	8	.4	.6	.3	-.011	.033	.105	.069	.009	.030	.096	.062	-.004	.019	.059	.129	.004	.018	.055	.135	.034
200	8	.8	.0	.0	-.002	.013	.038	.051	.000	.007	.020	.054	.000	.011	.034	.092	.000	.006	.019	.113	.033
200	8	.8	.0	.3	-.003	.019	.057	.060	.001	.011	.032	.055	-.001	.014	.043	.111	.000	.009	.027	.118	.033
200	8	.8	.6	.0	-.002	.014	.042	.052	.001	.007	.021	.058	-.001	.012	.035	.098	.001	.006	.019	.111	.033
200	8	.8	.6	.3	-.005	.021	.065	.066	.003	.013	.037	.060	-.002	.015	.046	.117	.001	.010	.029	.121	.035
800	4	.4	.0	.0	.000	.011	.032	.063	.001	.008	.024	.054	.000	.009	.027	.056	.000	.007	.021	.060	.050
800	4	.4	.0	.3	-.002	.016	.050	.061	.001	.011	.036	.054	.000	.012	.037	.052	.000	.009	.027	.058	.045
800	4	.4	.6	.0	-.001	.011	.036	.060	.001	.008	.024	.061	.000	.009	.028	.054	.000	.007	.021	.062	.051
800	4	.4	.6	.3	-.003	.018	.057	.054	.002	.012	.038	.062	.000	.012	.038	.054	.000	.009	.027	.058	.051
800	4	.8	.0	.0	-.001	.013	.039	.057	.000	.004	.013	.056	.000	.010	.029	.049	.000	.004	.011	.053	.053
800	4	.8	.0	.3	-.002	.016	.047	.059	.000	.005	.016	.057	.000	.011	.032	.052	.000	.004	.012	.052	.053
800	4	.8	.6	.0	-.001	.014	.040	.054	.000	.004	.013	.055	.000	.010	.029	.048	.000	.004	.011	.052	.052
800	4	.8	.6	.3	-.002	.016	.050	.057	.000	.005	.016	.059	.000	.011	.032	.052	.000	.004	.012	.048	.052
800	8	.4	.0	.0	.000	.007	.019	.050	.000	.006	.017	.061	.000	.005	.015	.063	.000	.005	.015	.073	.059
800	8	.4	.0	.3	-.001	.014	.041	.057	.001	.012	.037	.058	.000	.008	.023	.074	.000	.007	.022	.076	.058
800	8	.4	.6	.0	.000	.007	.022	.049	.001	.006	.018	.059	.000	.005	.016	.056	.000	.005	.014	.071	.052
800	8	.4	.6	.3	-.003	.016	.049	.056	.002	.013	.042	.058	.000	.008	.024	.070	.000	.007	.022	.076	.054
800	8	.8	.0	.0	.000	.007	.019	.048	.000	.003	.010	.063	.000	.005	.015	.055	.000	.003	.009	.074	.047
800	8	.8	.0	.3	-.001	.010	.029	.048	.000	.005	.016	.058	.000	.006	.019	.063	.000	.004	.012	.072	.053
800	8	.8	.6	.0	.000	.007	.021	.052	.000	.003	.010	.063	.000	.005	.016	.054	.000	.003	.009	.072	.049
800	8	.8	.6	.3	-.001	.011	.033	.052	.000	.006	.018	.051	.000	.006	.019	.065	.000	.004	.012	.069	.047

Table B3: Nonlinear GMM estimator using $w_i = y_{i,0}^1$ and $\mu_\lambda = 1$

Designs					GMM 1 step								GMM 2 step								
					α				β				α				β				J
N	T	α	ρ	δ	Bias	RMSE	qStd	Size	Bias	RMSE	qStd	Size	Bias	RMSE	qStd	Size	Bias	RMSE	qStd	Size	Size
200	4	.4	.0	.0	-.002	.026	.078	.060	.002	.022	.065	.056	-.001	.022	.068	.058	.002	.020	.059	.057	.047
200	4	.4	.0	.3	-.006	.042	.132	.058	.004	.033	.102	.055	-.002	.033	.098	.059	.002	.026	.079	.058	.044
200	4	.4	.6	.0	-.002	.026	.082	.057	.003	.021	.066	.052	-.002	.023	.070	.056	.002	.019	.058	.054	.047
200	4	.4	.6	.3	-.007	.044	.135	.064	.005	.032	.103	.052	-.002	.033	.102	.057	.003	.026	.078	.055	.045
200	4	.8	.0	.0	-.002	.032	.097	.064	.001	.011	.033	.053	-.001	.024	.074	.061	.001	.010	.029	.061	.044
200	4	.8	.0	.3	-.004	.039	.122	.068	.001	.014	.042	.058	-.004	.027	.087	.065	.001	.011	.034	.065	.044
200	4	.8	.6	.0	-.002	.032	.099	.072	.001	.011	.033	.060	-.002	.024	.077	.068	.001	.010	.029	.065	.045
200	4	.8	.6	.3	-.004	.039	.123	.072	.001	.013	.041	.064	-.003	.028	.089	.063	.001	.011	.034	.064	.042
200	8	.4	.0	.0	-.003	.015	.045	.054	.003	.015	.046	.055	-.001	.011	.035	.086	.001	.012	.035	.101	.036
200	8	.4	.0	.3	-.013	.040	.119	.066	.012	.039	.117	.070	-.004	.021	.060	.112	.004	.020	.058	.118	.031
200	8	.4	.6	.0	-.002	.015	.046	.057	.003	.015	.043	.061	-.001	.012	.037	.082	.001	.012	.034	.096	.037
200	8	.4	.6	.3	-.013	.040	.118	.069	.013	.037	.111	.071	-.005	.020	.060	.111	.005	.019	.057	.113	.032
200	8	.8	.0	.0	-.002	.013	.041	.051	.001	.007	.023	.058	-.001	.012	.034	.083	.001	.007	.020	.100	.035
200	8	.8	.0	.3	-.004	.021	.064	.058	.003	.013	.040	.061	-.002	.015	.044	.088	.001	.010	.029	.099	.035
200	8	.8	.6	.0	-.002	.014	.042	.052	.001	.007	.022	.056	-.001	.012	.035	.084	.001	.007	.020	.094	.033
200	8	.8	.6	.3	-.005	.022	.068	.063	.003	.014	.040	.058	-.003	.016	.045	.087	.002	.010	.030	.098	.034
800	4	.4	.0	.0	.000	.013	.039	.048	.000	.011	.033	.050	.000	.011	.033	.050	.000	.009	.029	.052	.046
800	4	.4	.0	.3	-.001	.020	.066	.048	.002	.016	.052	.049	.000	.016	.048	.054	.000	.012	.038	.055	.046
800	4	.4	.6	.0	.000	.013	.040	.050	.000	.010	.033	.048	.000	.011	.034	.053	.000	.009	.028	.048	.048
800	4	.4	.6	.3	-.001	.021	.067	.048	.002	.016	.051	.050	.000	.016	.049	.055	.001	.012	.037	.050	.048
800	4	.8	.0	.0	.000	.016	.049	.054	.000	.005	.017	.047	.000	.011	.036	.052	.000	.004	.014	.047	.045
800	4	.8	.0	.3	-.001	.019	.060	.062	.000	.006	.021	.056	.000	.013	.041	.050	.000	.005	.017	.051	.047
800	4	.8	.6	.0	.000	.016	.050	.058	.000	.005	.017	.049	-.001	.011	.035	.052	.000	.004	.014	.050	.043
800	4	.8	.6	.3	-.001	.020	.062	.063	.000	.007	.021	.056	.000	.013	.041	.054	.000	.005	.016	.050	.046
800	8	.4	.0	.0	-.001	.008	.022	.051	.001	.007	.023	.060	.000	.005	.016	.056	.000	.005	.016	.066	.051
800	8	.4	.0	.3	-.003	.020	.059	.059	.002	.019	.057	.059	-.001	.008	.025	.070	.001	.008	.024	.073	.051
800	8	.4	.6	.0	-.001	.008	.022	.047	.001	.007	.021	.063	.000	.006	.017	.053	.000	.005	.016	.070	.051
800	8	.4	.6	.3	-.005	.019	.058	.056	.004	.018	.055	.061	-.001	.008	.025	.068	.001	.008	.025	.075	.049
800	8	.8	.0	.0	-.001	.007	.020	.049	.000	.004	.011	.063	.000	.005	.016	.055	.000	.003	.009	.065	.048
800	8	.8	.0	.3	-.001	.011	.032	.055	.001	.007	.020	.057	.000	.006	.020	.060	.000	.004	.014	.067	.046
800	8	.8	.6	.0	-.001	.007	.021	.054	.000	.004	.011	.064	.000	.005	.016	.052	.000	.003	.009	.065	.046
800	8	.8	.6	.3	-.001	.011	.033	.050	.001	.007	.020	.054	-.001	.007	.020	.060	.000	.005	.013	.067	.044

Table B4: Nonlinear GMM estimator using $w_i = y_{i,0}^2$ and $\mu_\lambda = 0$

Designs					GMM 1 step								GMM 2 step								
					α				β				α				β				J
N	T	α	ρ	δ	Bias	RMSE	qStd	Size	Bias	RMSE	qStd	Size	Bias	RMSE	qStd	Size	Bias	RMSE	qStd	Size	Size
200	4	.4	.0	.0	-.024	.041	.121	.070	.014	.029	.087	.034	-.010	.027	.086	.133	.004	.018	.061	.107	.275
200	4	.4	.0	.3	-.046	.079	.244	.089	.023	.052	.174	.047	-.021	.042	.148	.183	.008	.027	.097	.129	.306
200	4	.4	.6	.0	-.060	.076	.207	.190	.050	.059	.163	.156	-.020	.036	.152	.241	.014	.025	.112	.196	.356
200	4	.4	.6	.3	-.130	.157	.383	.257	.097	.115	.316	.208	-.040	.054	.268	.314	.025	.038	.195	.258	.407
200	4	.8	.0	.0	-.019	.041	.137	.056	.005	.012	.040	.029	-.014	.032	.114	.159	.003	.010	.033	.107	.305
200	4	.8	.0	.3	-.027	.057	.211	.070	.005	.017	.059	.033	-.019	.037	.159	.186	.003	.012	.042	.121	.324
200	4	.8	.6	.0	-.039	.056	.179	.129	.015	.019	.060	.082	-.021	.036	.149	.227	.005	.011	.042	.140	.345
200	4	.8	.6	.3	-.058	.080	.269	.141	.018	.026	.094	.086	-.028	.044	.208	.243	.007	.014	.060	.157	.366
200	8	.4	.0	.0	-.023	.032	.089	.084	.025	.034	.096	.080	-.011	.019	.058	.194	.011	.018	.055	.186	.318
200	8	.4	.0	.3	-.117	.128	.270	.282	.102	.114	.275	.247	-.051	.059	.162	.434	.044	.052	.152	.407	.340
200	8	.4	.6	.0	-.085	.091	.182	.311	.103	.110	.244	.328	-.039	.042	.130	.446	.043	.046	.156	.479	.417
200	8	.4	.6	.3	-.358	.363	.475	.577	.362	.366	.527	.566	-.197	.198	.347	.685	.182	.184	.354	.684	.542
200	8	.8	.0	.0	-.025	.030	.085	.108	.009	.014	.044	.054	-.014	.019	.066	.226	.005	.010	.029	.160	.357
200	8	.8	.0	.3	-.056	.065	.200	.166	.021	.034	.123	.083	-.031	.036	.137	.322	.010	.018	.066	.227	.398
200	8	.8	.6	.0	-.056	.058	.136	.241	.033	.034	.091	.230	-.030	.033	.111	.369	.015	.017	.058	.360	.435
200	8	.8	.6	.3	-.137	.142	.333	.346	.089	.095	.256	.306	-.071	.073	.248	.496	.038	.042	.160	.466	.487
800	4	.4	.0	.0	-.020	.035	.104	.087	.011	.023	.075	.051	-.006	.016	.068	.198	.003	.011	.043	.147	.464
800	4	.4	.0	.3	-.043	.077	.232	.102	.019	.047	.163	.064	-.010	.026	.129	.235	.004	.018	.076	.182	.503
800	4	.4	.6	.0	-.058	.076	.200	.195	.051	.062	.160	.177	-.012	.022	.136	.292	.007	.017	.101	.244	.551
800	4	.4	.6	.3	-.138	.168	.386	.265	.098	.122	.315	.220	-.024	.038	.250	.347	.015	.028	.183	.300	.583
800	4	.8	.0	.0	-.017	.030	.100	.068	.003	.008	.029	.038	-.008	.018	.087	.204	.002	.006	.022	.142	.459
800	4	.8	.0	.3	-.019	.045	.179	.077	.003	.012	.048	.039	-.010	.023	.134	.229	.002	.007	.029	.150	.487
800	4	.8	.6	.0	-.043	.051	.159	.147	.014	.017	.052	.120	-.012	.023	.134	.261	.003	.007	.037	.215	.508
800	4	.8	.6	.3	-.059	.078	.254	.155	.018	.024	.088	.112	-.015	.030	.191	.285	.004	.009	.050	.223	.525
800	8	.4	.0	.0	-.023	.029	.077	.113	.024	.031	.083	.113	-.005	.010	.036	.189	.005	.009	.032	.184	.589
800	8	.4	.0	.3	-.121	.128	.268	.335	.107	.118	.273	.308	-.029	.034	.119	.412	.023	.029	.104	.384	.667
800	8	.4	.6	.0	-.086	.091	.176	.310	.105	.111	.234	.316	-.020	.024	.101	.384	.021	.024	.122	.394	.696
800	8	.4	.6	.3	-.368	.372	.496	.584	.371	.378	.545	.575	-.136	.136	.304	.622	.122	.125	.306	.619	.796
800	8	.8	.0	.0	-.022	.026	.073	.128	.008	.012	.037	.078	-.008	.011	.046	.240	.003	.005	.018	.161	.636
800	8	.8	.0	.3	-.056	.066	.192	.158	.018	.030	.122	.085	-.016	.021	.112	.293	.005	.010	.045	.201	.669
800	8	.8	.6	.0	-.052	.055	.126	.228	.031	.032	.086	.217	-.017	.019	.088	.345	.008	.009	.045	.322	.682
800	8	.8	.6	.3	-.143	.152	.334	.344	.085	.092	.256	.310	-.047	.050	.219	.454	.020	.024	.130	.418	.725

Table B5: Nonlinear GMM estimator using $w_i = y_{i,0}^0$ and $\mu_\lambda = 0$

Designs					GMM 1 step								GMM 2 step								
					α				β				α				β				J
N	T	α	ρ	δ	Bias	RMSE	qStd	Size	Bias	RMSE	qStd	Size	Bias	RMSE	qStd	Size	Bias	RMSE	qStd	Size	Size
200	4	.4	.0	.0	-.033	.050	.139	.100	.017	.035	.101	.041	-.015	.031	.101	.172	.006	.022	.069	.132	.423
200	4	.4	.0	.3	-.067	.101	.266	.136	.031	.067	.206	.080	-.030	.051	.169	.236	.012	.034	.111	.173	.457
200	4	.4	.6	.0	-.098	.108	.217	.281	.085	.092	.185	.250	-.039	.050	.169	.338	.026	.036	.131	.290	.519
200	4	.4	.6	.3	-.212	.227	.399	.364	.153	.170	.346	.320	-.083	.093	.287	.425	.053	.064	.221	.367	.564
200	4	.8	.0	.0	-.032	.048	.149	.076	.006	.014	.044	.040	-.024	.038	.130	.216	.004	.012	.038	.132	.456
200	4	.8	.0	.3	-.042	.070	.224	.094	.007	.021	.066	.041	-.031	.048	.176	.239	.005	.014	.050	.143	.483
200	4	.8	.6	.0	-.069	.078	.199	.174	.022	.025	.066	.129	-.035	.048	.170	.292	.009	.015	.053	.206	.513
200	4	.8	.6	.3	-.101	.112	.294	.191	.030	.036	.108	.139	-.050	.061	.235	.327	.013	.019	.074	.227	.530
200	8	.4	.0	.0	-.033	.041	.098	.143	.035	.043	.109	.137	-.015	.022	.063	.251	.016	.021	.061	.258	.416
200	8	.4	.0	.3	-.152	.162	.292	.381	.133	.149	.309	.351	-.072	.078	.172	.522	.060	.069	.162	.495	.447
200	8	.4	.6	.0	-.129	.132	.185	.466	.173	.175	.254	.518	-.066	.067	.133	.568	.084	.084	.160	.645	.555
200	8	.4	.6	.3	-.441	.443	.461	.711	.457	.461	.521	.703	-.284	.284	.332	.787	.276	.277	.352	.792	.683
200	8	.8	.0	.0	-.035	.040	.090	.176	.013	.018	.050	.094	-.021	.025	.070	.303	.007	.011	.032	.199	.501
200	8	.8	.0	.3	-.081	.088	.210	.242	.030	.045	.144	.141	-.045	.049	.149	.404	.014	.023	.076	.285	.532
200	8	.8	.6	.0	-.082	.083	.142	.354	.053	.054	.095	.368	-.049	.050	.112	.490	.026	.026	.062	.492	.578
200	8	.8	.6	.3	-.220	.221	.334	.511	.138	.139	.274	.489	-.137	.137	.251	.641	.073	.075	.173	.621	.643
800	4	.4	.0	.0	-.034	.049	.121	.149	.019	.031	.089	.073	-.010	.021	.081	.257	.004	.014	.052	.185	.618
800	4	.4	.0	.3	-.075	.101	.251	.171	.031	.064	.196	.107	-.021	.037	.148	.322	.008	.023	.095	.249	.662
800	4	.4	.6	.0	-.103	.116	.211	.308	.093	.101	.178	.293	-.024	.039	.155	.403	.017	.028	.121	.358	.708
800	4	.4	.6	.3	-.228	.244	.398	.382	.161	.178	.356	.345	-.071	.084	.276	.471	.041	.055	.219	.420	.734
800	4	.8	.0	.0	-.029	.041	.115	.113	.006	.011	.033	.047	-.015	.026	.107	.291	.003	.007	.027	.203	.620
800	4	.8	.0	.3	-.037	.059	.205	.111	.006	.016	.058	.055	-.019	.033	.154	.326	.003	.009	.039	.216	.642
800	4	.8	.6	.0	-.067	.077	.172	.224	.024	.026	.058	.197	-.024	.035	.152	.377	.006	.010	.044	.286	.667
800	4	.8	.6	.3	-.098	.110	.277	.235	.030	.035	.104	.196	-.033	.048	.213	.395	.008	.013	.065	.305	.686
800	8	.4	.0	.0	-.034	.040	.085	.228	.037	.044	.096	.242	-.009	.014	.041	.288	.010	.013	.038	.292	.767
800	8	.4	.0	.3	-.157	.164	.284	.497	.141	.154	.297	.444	-.053	.057	.130	.580	.041	.047	.117	.544	.831
800	8	.4	.6	.0	-.133	.137	.177	.507	.185	.187	.244	.563	-.049	.051	.111	.576	.062	.063	.129	.615	.878
800	8	.4	.6	.3	-.468	.469	.464	.753	.487	.489	.524	.744	-.251	.251	.306	.792	.241	.241	.313	.793	.920
800	8	.8	.0	.0	-.033	.036	.078	.252	.013	.016	.045	.163	-.014	.015	.051	.350	.005	.007	.020	.235	.812
800	8	.8	.0	.3	-.083	.090	.210	.278	.030	.044	.149	.181	-.032	.033	.126	.447	.010	.015	.056	.325	.837
800	8	.8	.6	.0	-.082	.083	.132	.404	.055	.055	.097	.442	-.037	.037	.096	.506	.018	.018	.050	.519	.851
800	8	.8	.6	.3	-.224	.226	.338	.544	.149	.153	.281	.537	-.113	.113	.225	.637	.059	.060	.149	.620	.875

Table B6: Nonlinear GMM estimator using $w_i = y_{i,0}^1$ and $\mu_\lambda = 0$

Designs					GMM 1 step								GMM 2 step								
					α				β				α				β				J
N	T	α	ρ	δ	Bias	RMSE	qStd	Size	Bias	RMSE	qStd	Size	Bias	RMSE	qStd	Size	Bias	RMSE	qStd	Size	Size
200	4	.4	.0	.0	-.002	.026	.078	.058	.001	.021	.065	.051	-.001	.023	.069	.062	.001	.020	.061	.062	.041
200	4	.4	.0	.3	-.005	.041	.125	.064	.002	.032	.097	.049	-.001	.033	.099	.063	.001	.027	.080	.063	.041
200	4	.4	.6	.0	-.003	.026	.081	.065	.002	.021	.065	.053	-.001	.024	.071	.063	.001	.020	.060	.059	.044
200	4	.4	.6	.3	-.006	.042	.132	.064	.005	.031	.100	.050	-.002	.034	.101	.059	.002	.026	.079	.061	.044
200	4	.8	.0	.0	-.003	.032	.099	.062	.001	.011	.034	.058	-.002	.025	.074	.060	.001	.010	.030	.063	.048
200	4	.8	.0	.3	-.005	.040	.120	.068	.001	.013	.044	.065	-.003	.028	.088	.067	.001	.012	.036	.067	.044
200	4	.8	.6	.0	-.004	.032	.100	.074	.001	.011	.035	.058	-.003	.025	.076	.065	.001	.010	.030	.069	.042
200	4	.8	.6	.3	-.006	.041	.126	.075	.002	.014	.044	.069	-.004	.028	.089	.069	.001	.012	.035	.064	.046
200	8	.4	.0	.0	-.003	.014	.042	.056	.002	.014	.041	.057	-.001	.012	.036	.088	.001	.012	.034	.092	.037
200	8	.4	.0	.3	-.011	.035	.101	.068	.009	.032	.098	.065	-.003	.019	.059	.105	.004	.019	.055	.106	.037
200	8	.4	.6	.0	-.003	.015	.045	.059	.003	.014	.041	.061	-.001	.012	.037	.090	.001	.012	.034	.092	.035
200	8	.4	.6	.3	-.015	.037	.110	.077	.014	.036	.103	.076	-.005	.021	.061	.121	.006	.020	.056	.112	.038
200	8	.8	.0	.0	-.002	.013	.040	.056	.001	.007	.022	.060	-.001	.012	.035	.089	.001	.007	.021	.089	.036
200	8	.8	.0	.3	-.005	.020	.059	.058	.002	.013	.036	.055	-.002	.015	.044	.097	.001	.010	.030	.086	.039
200	8	.8	.6	.0	-.002	.013	.041	.059	.001	.007	.022	.057	-.001	.012	.036	.084	.001	.007	.021	.090	.036
200	8	.8	.6	.3	-.005	.021	.065	.060	.003	.013	.039	.063	-.002	.015	.045	.090	.002	.010	.029	.086	.036
800	4	.4	.0	.0	.000	.012	.038	.051	.000	.010	.032	.051	.000	.011	.032	.059	.000	.009	.029	.052	.046
800	4	.4	.0	.3	-.001	.019	.062	.052	.001	.016	.048	.052	.000	.015	.047	.054	.000	.012	.038	.051	.048
800	4	.4	.6	.0	.000	.013	.039	.053	.000	.010	.033	.051	.000	.011	.034	.054	.000	.009	.029	.051	.048
800	4	.4	.6	.3	-.002	.020	.064	.046	.001	.016	.049	.051	-.001	.015	.047	.055	.000	.012	.038	.050	.046
800	4	.8	.0	.0	-.001	.016	.048	.050	.000	.005	.017	.050	.000	.012	.035	.053	.000	.005	.014	.051	.046
800	4	.8	.0	.3	-.001	.019	.060	.060	.000	.007	.021	.058	.000	.014	.041	.051	.000	.005	.017	.054	.047
800	4	.8	.6	.0	-.002	.016	.050	.055	.000	.005	.017	.051	-.001	.012	.036	.051	.000	.005	.014	.055	.046
800	4	.8	.6	.3	-.002	.020	.063	.066	.000	.007	.022	.067	-.001	.013	.041	.055	.000	.005	.016	.061	.049
800	8	.4	.0	.0	-.001	.007	.021	.049	.001	.007	.021	.062	.000	.005	.016	.060	.001	.005	.016	.069	.052
800	8	.4	.0	.3	-.003	.016	.051	.061	.003	.016	.049	.060	.000	.008	.025	.069	.001	.008	.025	.071	.052
800	8	.4	.6	.0	-.001	.007	.022	.049	.001	.007	.020	.062	.000	.006	.017	.054	.000	.005	.016	.070	.052
800	8	.4	.6	.3	-.005	.018	.054	.060	.005	.017	.051	.060	-.001	.008	.026	.071	.001	.008	.025	.073	.051
800	8	.8	.0	.0	-.001	.007	.020	.053	.000	.003	.011	.058	.000	.005	.015	.056	.000	.003	.010	.063	.048
800	8	.8	.0	.3	-.001	.010	.030	.053	.001	.006	.018	.055	-.001	.006	.019	.062	.000	.005	.014	.071	.051
800	8	.8	.6	.0	-.001	.007	.021	.054	.000	.004	.011	.057	.000	.006	.016	.054	.000	.003	.009	.066	.048
800	8	.8	.6	.3	-.002	.011	.032	.053	.001	.006	.019	.054	-.001	.007	.020	.061	.000	.005	.014	.070	.046

Table B7: Linearized GMM estimator using $w_i = y_{i,0}^2$ and $\mu_\lambda = 1$

Designs					GMM 1 step								GMM 2 step								
					α				β				α				β				J
N	T	α	ρ	δ	Bias	RMSE	qStd	Size	Bias	RMSE	qStd	Size	Bias	RMSE	qStd	Size	Bias	RMSE	qStd	Size	Size
200	8	.4	.0	.0	-.006	.023	.071	.021	.005	.020	.062	.018	-.005	.020	.058	.021	.003	.018	.055	.024	.016
200	8	.4	.0	.3	-.030	.048	.158	.065	.018	.045	.148	.043	-.017	.033	.101	.094	.003	.031	.099	.063	.026
200	8	.4	.6	.0	-.007	.026	.080	.024	.007	.020	.062	.022	-.005	.021	.061	.020	.003	.018	.054	.024	.018
200	8	.4	.6	.3	-.036	.054	.179	.071	.020	.046	.151	.046	-.020	.034	.105	.099	.004	.031	.099	.069	.029
200	8	.8	.0	.0	-.004	.028	.095	.017	.001	.014	.043	.022	-.009	.025	.082	.025	.000	.013	.040	.030	.017
200	8	.8	.0	.3	-.019	.045	.163	.029	.004	.022	.080	.024	-.024	.036	.138	.054	-.003	.021	.070	.038	.024
200	8	.8	.6	.0	-.004	.029	.103	.019	.001	.014	.043	.026	-.011	.026	.090	.038	.000	.013	.039	.034	.018
200	8	.8	.6	.3	-.020	.045	.157	.030	.006	.025	.081	.022	-.021	.033	.117	.037	.000	.021	.069	.025	.015
800	8	.4	.0	.0	-.001	.012	.035	.017	.001	.010	.029	.019	-.001	.009	.028	.022	.000	.009	.026	.026	.023
800	8	.4	.0	.3	-.010	.026	.089	.051	.005	.022	.077	.043	-.005	.016	.049	.062	-.001	.015	.048	.043	.040
800	8	.4	.6	.0	-.001	.013	.039	.021	.001	.010	.030	.018	-.001	.009	.029	.023	.000	.009	.025	.022	.024
800	8	.4	.6	.3	-.011	.028	.094	.056	.005	.022	.078	.050	-.005	.015	.048	.069	-.001	.015	.047	.040	.039
800	8	.8	.0	.0	-.001	.016	.055	.016	.000	.007	.023	.025	-.003	.013	.045	.035	.000	.006	.020	.040	.024
800	8	.8	.0	.3	-.008	.025	.106	.022	.000	.012	.039	.020	-.008	.018	.072	.049	-.002	.010	.034	.033	.031
800	8	.8	.6	.0	-.001	.017	.062	.022	.000	.007	.023	.034	-.002	.013	.048	.043	-.001	.006	.019	.044	.026
800	8	.8	.6	.3	-.008	.023	.089	.017	.001	.012	.040	.020	-.006	.015	.055	.027	-.001	.009	.032	.019	.021

Table B8: Linearized GMM estimator using $w_i = y_{i,0}^0$ and $\mu_\lambda = 1$

Designs					GMM 1 step								GMM 2 step								
					α				β				α				β				J
N	T	α	ρ	δ	Bias	RMSE	qStd	Size	Bias	RMSE	qStd	Size	Bias	RMSE	qStd	Size	Bias	RMSE	qStd	Size	Size
200	8	.4	.0	.0	-.004	.022	.069	.024	.003	.018	.056	.018	-.004	.019	.059	.025	.002	.017	.052	.026	.020
200	8	.4	.0	.3	-.014	.039	.121	.046	.004	.036	.123	.035	-.010	.029	.088	.064	-.001	.029	.091	.053	.026
200	8	.4	.6	.0	-.005	.025	.076	.024	.004	.019	.059	.019	-.005	.020	.061	.023	.003	.017	.053	.027	.018
200	8	.4	.6	.3	-.028	.050	.160	.065	.015	.044	.141	.048	-.016	.031	.094	.082	.002	.030	.093	.060	.029
200	8	.8	.0	.0	-.005	.028	.094	.027	.001	.012	.037	.021	-.011	.025	.084	.046	.000	.012	.036	.033	.015
200	8	.8	.0	.3	-.010	.037	.114	.020	.002	.021	.065	.014	-.015	.030	.095	.028	-.001	.019	.057	.021	.015
200	8	.8	.6	.0	-.005	.029	.099	.022	.002	.012	.037	.023	-.010	.025	.087	.044	.000	.012	.035	.034	.016
200	8	.8	.6	.3	-.014	.040	.131	.019	.005	.022	.071	.016	-.017	.030	.099	.030	.000	.019	.059	.020	.018
800	8	.4	.0	.0	.000	.011	.035	.018	.001	.010	.031	.017	-.001	.009	.028	.019	.000	.009	.027	.023	.017
800	8	.4	.0	.3	-.004	.020	.063	.045	.002	.017	.055	.034	-.003	.014	.042	.059	-.001	.014	.044	.048	.039
800	8	.4	.6	.0	.000	.013	.038	.017	.001	.011	.033	.016	-.001	.010	.029	.018	.000	.009	.027	.025	.020
800	8	.4	.6	.3	-.010	.025	.083	.058	.003	.019	.066	.039	-.004	.015	.044	.061	-.001	.015	.045	.046	.037
800	8	.8	.0	.0	-.001	.016	.053	.037	.000	.006	.019	.036	-.003	.012	.042	.049	.000	.006	.017	.047	.033
800	8	.8	.0	.3	-.002	.018	.059	.012	.000	.011	.033	.017	-.004	.014	.044	.015	.000	.009	.028	.020	.020
800	8	.8	.6	.0	-.001	.016	.057	.033	.000	.006	.019	.039	-.003	.013	.043	.046	.000	.006	.017	.047	.033
800	8	.8	.6	.3	-.003	.021	.067	.013	.001	.011	.037	.016	-.004	.014	.045	.012	.000	.009	.028	.021	.019

Table B9: Linearized GMM estimator using $w_i = y_{i,0}^1$ and $\mu_\lambda = 1$

Designs					GMM 1 step								GMM 2 step								
					α				β				α				β				J
N	T	α	ρ	δ	Bias	RMSE	qStd	Size	Bias	RMSE	qStd	Size	Bias	RMSE	qStd	Size	Bias	RMSE	qStd	Size	Size
200	8	.4	.0	.0	-.006	.026	.085	.022	.005	.025	.077	.021	-.004	.022	.066	.024	.003	.022	.064	.019	.019
200	8	.4	.0	.3	-.034	.062	.200	.053	.022	.057	.201	.043	-.019	.039	.120	.091	.004	.038	.124	.060	.030
200	8	.4	.6	.0	-.007	.029	.089	.026	.005	.025	.075	.021	-.005	.022	.068	.024	.003	.021	.063	.019	.020
200	8	.4	.6	.3	-.038	.063	.209	.066	.023	.055	.184	.050	-.021	.039	.118	.090	.004	.038	.119	.057	.032
200	8	.8	.0	.0	-.003	.030	.102	.015	.002	.017	.054	.016	-.010	.027	.087	.022	.001	.015	.049	.024	.020
200	8	.8	.0	.3	-.021	.050	.195	.025	.004	.028	.100	.030	-.023	.040	.156	.061	-.002	.025	.088	.042	.028
200	8	.8	.6	.0	-.005	.032	.111	.019	.002	.017	.054	.022	-.011	.028	.097	.037	.000	.015	.050	.028	.019
200	8	.8	.6	.3	-.018	.047	.169	.022	.003	.028	.098	.021	-.017	.037	.126	.031	-.001	.024	.077	.026	.020
800	8	.4	.0	.0	-.002	.013	.040	.017	.002	.012	.038	.018	-.001	.010	.032	.023	.000	.011	.033	.025	.019
800	8	.4	.0	.3	-.011	.030	.104	.046	.008	.028	.105	.043	-.005	.018	.058	.055	.001	.019	.064	.048	.046
800	8	.4	.6	.0	-.001	.014	.042	.015	.002	.012	.037	.018	-.001	.011	.033	.024	.000	.010	.032	.032	.025
800	8	.4	.6	.3	-.011	.030	.101	.052	.008	.027	.095	.049	-.005	.018	.056	.057	.001	.018	.060	.048	.043
800	8	.8	.0	.0	-.001	.017	.061	.012	.000	.009	.029	.020	-.002	.014	.048	.031	.000	.008	.027	.028	.023
800	8	.8	.0	.3	-.007	.027	.116	.025	.000	.014	.053	.025	-.008	.019	.075	.053	-.001	.012	.043	.044	.037
800	8	.8	.6	.0	-.002	.018	.068	.021	.000	.009	.029	.025	-.002	.014	.051	.042	.000	.007	.026	.034	.029
800	8	.8	.6	.3	-.006	.024	.095	.011	.001	.014	.047	.021	-.006	.017	.063	.032	-.001	.011	.039	.026	.020