A Simple Estimator for Short Panels with Common Factors☆

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

[a] *Faculty of Economics and Business, University of Groningen.*
[b] *Department of Econometrics and Business Statistics, Monash University.*

Abstract

We develop a new method 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 dependent and independent variables of the model. Not only is the proposed approach easy to implement relative to existing ones, but since the model is effectively parameterized in a more parsimonious way the resulting estimator can be asymptotically more efficient. 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.

1. Introduction

The panel two way error components model is very popular in social sciences because it offers greater scope for controlling for unobservables compared to time series or cross-sectional models alone. However, from at least as far back as Holtz-Eakin et al. (1988)

☆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

it has been pointed out that this structure is potentially too restrictive since it assumes that the error components enter in an additive fashion. Consequently, there is by now a substantial literature on estimating panel data models with common factors, which allow multiple error components to enter in a multiplicative fashion. For the case where T is fixed, where T denotes the number of time series observations in the panel, methods that deal with common factors have been proposed by Nauges and Thomas (2003a), Ahn et al. (2001), Robertson and Sarafidis (2015), Bai (2013), Hayakawa (2012) and Hayakawa et al. (2014), among others.[1]

Despite this progress, the above methods may however suffer from certain shortcomings. In particular, since the factor component enters into the model multiplicatively, the computational burden involved can be substantial. In addition, there is often no underlying theory to guide the selection of good starting values for a potentially large number of nuisance parameters, which is a requirement for the majority of these methods. Local minima-related problems might also arise, depending on the values of the unobserved factors; see e.g. Juodis and Sarafidis (2015) for a detailed discussion on this specific problem.

These issues might explain why the aforementioned procedures appear to have remained largely unused by empirical practitioners, despite their theoretical appeal. In this paper we propose a simple estimation method that involves substituting the unknown factors with time-specific weighted averages of the dependent and independent variables of the model. The remaining parameters are estimated using the method of moments, or non-linear least squares. The gains of such a strategy are threefold. First, the resulting estimation procedure is substantially easier to implement because the unobserved factors are replaced by observed data. Second, the issue of how to initialize nuisance parameters is reduced to the rather more trivial task of performing a grid search over structural parameter values. Finally, the estimation method can be linearized in a straightforward way, leading to a closed form solution of the optimization problem, which can provide a consistent starting value of the structural parameters for the non-linear estimation methods.

We remark that since our model is effectively parameterized in a more parsimo-

[1] We refer the interested reader to Section 5 in Sarafidis and Wansbeek (2012) for an overview of this literature and Juodis and Sarafidis (2015) for a more recent in-depth analysis.

nious way, the resulting estimator can be asymptotically more efficient than existing ones. Furthermore, our methodology has the advantage that it is easily extendable to allow common *observed* factors, and it can accommodate unbalanced panels in a straightforward way.

Our method resembles intuitively the approach employed by the Common Correlated Effects (CCE) estimator developed 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 T large, whereas here the focus is on panels with fixed T and endogeneity is allowed.

Using simulated data, we show that the proposed estimators perform very well in finite samples. 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 active ongoing research, not only among economists but also across international environmental agencies, regulators, water utilities and the general public.

2. Contribution

2.1. Model and Assumptions

We consider the following dynamic panel data model with a multi-factor error structure:

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

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 assumed to bounded in absolute value 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:

$$\boldsymbol{y}_i = \alpha \boldsymbol{y}_{i,-1} + \sum_{k=1}^{K} \beta_k \boldsymbol{x}_i^{(k)} + \boldsymbol{F} \boldsymbol{\lambda}_i + \boldsymbol{\varepsilon}_i, \quad i = 1, \ldots, N, \tag{2}$$

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 $\tau(t,1), \ldots, \tau(t,K)$.[2]

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.[3] 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 even the less restrictive condition $\mathrm{E}\left(\varepsilon_{i,t} | \boldsymbol{\lambda}_i', \boldsymbol{x}_{i,0:\tau(t,1)}^{(1)}, \ldots, \boldsymbol{x}_{i,0:\tau(t,K)}^{(K)}\right) = 0$. 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 $\varepsilon_{i,t}$

[2] Shorthand notation $\boldsymbol{x}_{i,s:q}, s \leq q$ is used to denote the vectors of the form $\boldsymbol{x}_{i,s:q} = (x_{i,s}, \ldots, x_{i,q})'$.

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

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.

2.2. Method and Main Results

Suppose that there exists an $[L \times 1]$ vector of individual-specific weights $\boldsymbol{w}_i$ with finite second moments 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 (2) by $\boldsymbol{w}_i'$ and taking expectations we obtain

$$\boldsymbol{M}_y = \alpha \boldsymbol{M}_{y,-1} + \sum_{k=1}^{K} \beta_k \boldsymbol{M}_x^{(k)} + \boldsymbol{F}\boldsymbol{G}_w, \tag{3}$$

where $\boldsymbol{M}_y = \mathrm{E}\left(\boldsymbol{y}_i \boldsymbol{w}_i'\right)$, and so on for the remaining "M" (moments) matrices, while $\boldsymbol{G}_w = \mathrm{E}\left(\boldsymbol{\lambda}_i \boldsymbol{w}_i'\right)$ 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

$$\begin{aligned}
\boldsymbol{y}_i &= \alpha \boldsymbol{y}_{i,-1} + \sum_{k=1}^{K} \beta_k \boldsymbol{x}_i^{(k)} + \left(\boldsymbol{M}_y - \alpha \boldsymbol{M}_{y,-1} - \sum_{k=1}^{K} \beta_k \boldsymbol{M}_x^{(k)}\right) \boldsymbol{\lambda}_i + \boldsymbol{\varepsilon}_i \\
&= \alpha \boldsymbol{y}_{i,-1} + \sum_{k=1}^{K} \beta_k \boldsymbol{x}_i^{(k)} + \left(\boldsymbol{I}_T \otimes \boldsymbol{\lambda}_i'\right) \mathrm{vec}\left(\boldsymbol{M}_y' - \alpha \boldsymbol{M}_{y,-1}' - \sum_{k=1}^{K} \beta_k \left(\boldsymbol{M}_x^{(k)}\right)'\right) + \boldsymbol{\varepsilon}_i,
\end{aligned} \tag{4}$$

where, to save on notation, $\boldsymbol{\lambda}_i$ is reparametrized in terms of the original vector of factor loadings multiplied by $\boldsymbol{G}_w^{-1}$, i.e. $\boldsymbol{\lambda}_i \equiv \boldsymbol{G}_w^{-1} \boldsymbol{\lambda}_i$, while $\mathrm{vec}(\cdot)$ denotes the column stacking operator.

Let $\boldsymbol{z}_i = \left(\boldsymbol{y}'_{i,-1}, \left(\boldsymbol{x}^{(1)}_{i,0:\tau(T,1)}\right)', \ldots, \left(\boldsymbol{x}^{(K)}_{i,0:\tau(T,K)}\right)'\right)'$ 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. 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 ζ_t valid instruments at time t from the vector $\boldsymbol{z}_i$. As a result, $\boldsymbol{S}$ has dimension $[\zeta \times dT]$, where $\zeta \equiv \sum_{t=1}^{T} \zeta_t$. Define $\boldsymbol{Z}'_i \equiv \boldsymbol{S}(\boldsymbol{I}_T \otimes \boldsymbol{z}_i)$. Then under Assumptions 1-2, the following set of population moment conditions is valid by construction:

$$\mathrm{E}(\boldsymbol{Z}'_i \boldsymbol{\varepsilon}_i) = \boldsymbol{S}\,\mathrm{E}(\operatorname{vec}(\boldsymbol{z}_i \boldsymbol{\varepsilon}'_i)) = \boldsymbol{0}_\zeta. \tag{5}$$

Thus, pre-multiplying (4) by $\boldsymbol{Z}'_i$ and taking expectations yields

$$\begin{aligned} \boldsymbol{m} &= \alpha \boldsymbol{m}_{-1} + \sum_{k=1}^{K} \beta_k \boldsymbol{m}_k + \boldsymbol{S}(\boldsymbol{I}_T \otimes \boldsymbol{G}) \operatorname{vec}\left(\boldsymbol{M}'_y - \alpha \boldsymbol{M}'_{y,-1} - \sum_{k=1}^{K} \beta_k \left(\boldsymbol{M}^{(k)}_x\right)'\right) \\ &= \alpha \boldsymbol{m}_{-1} + \sum_{k=1}^{K} \beta_k \boldsymbol{m}_k + \boldsymbol{S} \operatorname{vec}\left(\boldsymbol{G}\boldsymbol{M}'_y - \alpha \boldsymbol{G}\boldsymbol{M}'_{y,-1} - \sum_{k=1}^{K} \beta_k \boldsymbol{G} \left(\boldsymbol{M}^{(k)}_x\right)'\right), \end{aligned} \tag{6}$$

where $\boldsymbol{G} = \mathrm{E}(\boldsymbol{z}_i \boldsymbol{\lambda}'_i)$ is a $[d \times L]$ matrix that contains nuisance parameters, or

$$\boldsymbol{m} = \alpha \boldsymbol{m}_{-1} + \sum_{k=1}^{K} \beta_k \boldsymbol{m}_k + \boldsymbol{S}\left[\left(\boldsymbol{M}_y - \alpha \boldsymbol{M}_{y,-1} - \sum_{k=1}^{K} \beta_k \boldsymbol{M}^{(k)}_x\right) \otimes \boldsymbol{I}_d\right] \boldsymbol{g}, \tag{7}$$

with $\boldsymbol{g} = \operatorname{vec}(\boldsymbol{G})$.

Since $\boldsymbol{F}$ has been replaced by observed data, starting values for the "structural" parameters provide a close form solution for a set of starting values for $\boldsymbol{g}$. As a result, estimation can be implemented initially using an iterative procedure, where given $(\alpha, \beta_1, \ldots, \beta_K)$ we estimate $\boldsymbol{g}$, and given $\boldsymbol{g}$ we estimate $(\alpha, \beta_1, \ldots, \beta_K)$. Subsequently, a grid search over the structural parameters can be performed to find the minimum. This procedure is very simple.

Remark 1. The vector of estimating equations in (6) can be easily linearized by defining the new parameters $\boldsymbol{G}_0 \equiv -\alpha \boldsymbol{G}$, $\boldsymbol{G}_k \equiv -\beta_k \boldsymbol{G}$. In this case the total number of parameters in (6) is given by

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

Compared to (6), 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. More details regarding the linearized estimator can be found in the Online Supplementary Appendix of this paper.

As the vector of estimating equations in (7) 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. In particular, it can be shown that

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

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 not necessarily strictly exogenous. As a result, $\boldsymbol{G}$ can be identified only in a block triangular fashion as well. To see this consider equation (1) 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 $t = T$:

$$\mathrm{E}[y_{i,T-1}\boldsymbol{\lambda}_i'], \mathrm{E}[x_{i,T}^{(1)}\boldsymbol{\lambda}_i'], \ldots, \mathrm{E}[x_{i,T}^{(K)}\boldsymbol{\lambda}_i']. \tag{8}$$

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 $\boldsymbol{g}$, that is, identification of $\boldsymbol{g}$ is up to a normalization. As a result, following Juodis and Sarafidis (2015) one can conclude that

$$\xi(L, \tau(T,1), \ldots, \tau(T,K)) = (K+1)\frac{L(L-1)}{2}. \tag{9}$$

Notice that the expression above remains constant even if one assumes that the covariates are endogenous, i.e. $\tau(t,k) = t - 1$.

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

$$\boldsymbol{\mu}_i(\boldsymbol{\theta}) = \boldsymbol{Z}_i' \left(\boldsymbol{y}_i - \alpha \boldsymbol{y}_{i,-1} - \sum_{k=1}^{K} \beta_k \boldsymbol{x}_i^{(k)} \right) - \boldsymbol{S} \left[\left(\left(\boldsymbol{y}_i - \alpha \boldsymbol{y}_{i,-1} - \sum_{k=1}^{K} \beta_k \boldsymbol{x}_i^{(k)} \right) \boldsymbol{w}_i' \right) \otimes \boldsymbol{I}_d \right] \widetilde{\boldsymbol{g}}, \tag{10}$$

where $\widetilde{\boldsymbol{g}} = \left(\boldsymbol{g}_r', (\boldsymbol{g} \backslash \boldsymbol{g}_r)' \right)'$ 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. *Define* $\boldsymbol{\mu}_N(\boldsymbol{\theta}) = N^{-1} \sum_{i=1}^{N} \boldsymbol{\mu}_i(\boldsymbol{\theta})$ *and let*

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

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. *This is straightforward enough; see e.g. Newey and McFadden (1994) for details.*

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 (5). The same result applies if one replaces $\boldsymbol{\Delta}$ by a consistent estimate.

Remark 2. 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.[4] In particular, the GMM estimator of 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, in practice the structural parameters of the model can be estimated based on the simple procedure described below (7).

Remark 3. 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 (2015).

2.3. Observed Factors

Often the empirical practitioner may wish to include variables that are common across individual entities, such as interest rates, unemployment rate etc., which can be viewed as common *observed* factors. It turns out that the extension of our estimator to such situation is trivial. To illustrate, consider the following model in vector form

$$\boldsymbol{y}_i = \alpha \boldsymbol{y}_{i,-1} + \sum_{k=1}^{K} \beta_k \boldsymbol{x}_i^{(k)} + \boldsymbol{F}^o \boldsymbol{\lambda}_i^o + \boldsymbol{F}^u \boldsymbol{\lambda}_i^u + \boldsymbol{\varepsilon}_i, \tag{11}$$

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

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

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

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

[4] See Theorem 3 in their paper.

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

$$\begin{aligned}\boldsymbol{y}_i &= \alpha \boldsymbol{y}_{i,-1} + \sum_{k=1}^{K} \beta_k \boldsymbol{x}_i^{(k)} + \left(\boldsymbol{I}_T \otimes \widetilde{\boldsymbol{\lambda}}_i^o\right)' \operatorname{vec}(\boldsymbol{F}^o) \\ &\quad + \left(\boldsymbol{I}_T \otimes \widetilde{\boldsymbol{\lambda}}_i^u\right)' \operatorname{vec}\left(\boldsymbol{M}_y - \alpha \boldsymbol{M}_{y,-1} - \sum_{k=1}^{K} \beta_k \boldsymbol{M}_x^{(k)}\right) + \boldsymbol{\varepsilon}_i. \end{aligned} \tag{12}$$

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

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

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

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

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

Finally, analogously to the model without observed factors, 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.

3. Implementation

3.1. Choices for $\boldsymbol{w}_i$

For ease of exposition and w.l.o.g. let $L = 1$. In this case the factor component can be written as $\lambda_i f_t$. A specific choice for the weights is to simply set $w_i = 1$, such that $\boldsymbol{M}_y$ ($\boldsymbol{M}_x^k$) in (3) becomes the crude time-specific average of $\boldsymbol{y}_i$ ($\boldsymbol{x}_i^k$). This particular strategy resembles the CCE estimator of Pesaran (2006), as the model is augmented by cross-sectional averages of the dependent and independent variables. The choice requires that $\mathrm{E}(\lambda_i) \neq 0$, which can be a plausible condition in several applications.

Juodis and Sarafidis (2015) provide finite sample evidence, showing 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 4. The resulting estimator is asymptotically more efficient than the so-called FIVU estimator proposed by Robertson and Sarafidis (2015), which employs the same estimating equations given in (6) but with the term inside the vec operator replaced by $\boldsymbol{F}$. The reason is that the proposed estimator requires estimating $T+1$ parameters, while FIVU estimates $1+2T-1$ parameters. To put it differently, the former estimator utilizes T additional moment conditions, i.e. those in equation (3). Of course, these become redundant when $\boldsymbol{G}_w^{-1}$ does not exist. 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.

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 requires $g_0 \neq 0$. This will hold true unless the initial observation is uncorrelated with λ_i, which is in general an implausible condition that violates the "fixed effects framework".[5]

Remark 5. 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 (6) 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 normalized by g_0.

Alternative choices of w_i can be obtained based on $y_{i,s}$, $s > 0$, or on powers of $y_{i,0}$ or $x_{i,0}$, such as $y_{i,0}^2$ and so on. The latter strategy does not require the distribution of $y_{i,t}$ to be symmetric because non-central moments are used. Notice that the condition (say) $\mathrm{E}\left(y_{i,0}^2 y_{i,t}\right) \neq 0$ is easy to check by replacing the expectation with the sample average $N^{-1}\sum_{i=1}^{N} w_i y_{i,t}$. The next section investigates the finite sample performance of the estimator that makes use of $w_i \in \{1, y_{i,0}, y_{i,0}^2\}$.

[5]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.

Our approach 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.

3.2. Testing for the Rank of $\boldsymbol{G}_w$

The full rank assumption on $\boldsymbol{G}_w$ can be verified in two different ways. First of all, it is straightforward to see that lack of invertibility of $\boldsymbol{G}_w$ implies that the factor component is not spanned by weighted averages of the observed data. As a result, the model becomes mis-specified and thereby the moment conditions are violated. This violation will be reflected in the overidentifying restrictions test statistic that is readily available within our procedure.

Secondly, the full rank assumption on $\boldsymbol{G}_w$ can be verified based on an examination of the rank of a matrix that contains the observed data. In particular, define the following matrix

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

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.* $\operatorname{rk}(\boldsymbol{M}) = (K+2)L$. *Then $\boldsymbol{G}_w$ is invertible.*

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

As a result of the theorem above, one could in principle verify prior to estimation if the unobserved matrix $\boldsymbol{G}_w$ has full rank by checking whether the $[T \times (K+2)L]$ matrix $\boldsymbol{M}$ of observed data satisfies $\operatorname{rk}(\boldsymbol{M}) = (K+2)L$. 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.[6] In the Monte Carlo section

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

that follows, we investigate the properties of the test statistic proposed by Robin and Smith (2000), 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, available in the Online Supplementary Appendix of the paper, demonstrate that the method works very well in our set up.

Remark 6. The rank condition on $\boldsymbol{G}_w$ 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.

3.3. 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 (Bayesian) Information Criterion (BIC):

$$S_N\left(L\right) = N \times Q_N\left(\widehat{\boldsymbol{\theta}}\left(\boldsymbol{\Omega}_N\right)|L\right) - \ln\left(N\right) \times h\left(L\right), \tag{13}$$

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

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

Proposition 2 follows directly from Theorem 5 in Robertson and Sarafidis (2015), with the only difference being that $\boldsymbol{F}$ in their model does not enter the objective function in the present case. Hence we do not replicate the proof here; the interested reader may refer to that paper.

Proposition 2 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. Finite Sample Evidence

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

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

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

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

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

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

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

$$\varepsilon_{i,t}^{y} \sim \mathcal{N}\left(0,1\right); \quad \varepsilon_{i,t}^{x} \sim \mathcal{N}\left(0,\sigma_x^2\right), \quad t \geq 0.$$

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

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

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

$$x_{i,0} = \sum_{\ell=1}^{L}\gamma_{\ell,i} + \varepsilon_{i,0}^{x}; \quad y_{i,0} = \sum_{\ell=1}^{L}\lambda_{\ell,i} + \varepsilon_{i,0}^{y},$$

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 remains always 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

We mainly consider three non-linear estimators based on different choices for w_i. In particular, NC(h) and NY(h) denote the non-linear GMM estimators that make use of $w_i = 1$ and $w_i = y_{i,0}$ respectively, while NY2(h) is the corresponding estimator making use of $w_i = y_{i,0}^2$. The argument $h \in \{1, 2\}$ 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 Appendix B 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, with nominal size being equal to 5%. For all two step estimators we report results using corrected standard errors.[7] In addition, for the two step GMM estimators we also report the empirical size of the overidentifying restrictions (J) test statistic.

Tables B1 and B2 present results for NC($\cdot$) and NY($\cdot$), respectively. In each case the top (bottom) panel corresponds to $\mu_\lambda = 1$ ($\mu_\lambda = 0$). For $\mu_\lambda = 1$, NC($\cdot$) performs very well in that there is 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 $\mu_\lambda = 0$ NC($\cdot$) is biased and size-distorted, as expected. However, the J test statistic appears to have good power in picking this up. In summary, NC($\cdot$) performs well when it is consistent and the J test statistic has good power to identify cases where it is not.

In regards to NY($\cdot$), the performance of the estimator is largely unaffected by the value of μ_λ, as expected. 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 NC($\cdot$), as reflected by the larger values of RMSE and qStd, because it makes use of T less moment conditions.

The performance of NY2($\cdot$) is very similar to NC($\cdot$) and therefore we do not discuss this separately. The results can been found in the Online Supplementary Appendix. The appendix also contains results for the case $\rho = 0$, as well as for the performance of the linearized version of all estimators, which remains satisfactory albeit it is "pound-for-pound" inferior to the performance of the non-linear estimators. Therefore it appears that extra simplicity comes with a cost.

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

[7] The derivations are based on Windmeijer (2005) and are documented in the Online Supplementary Appendix of the paper.

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.[8]

Urban water networks are characterised by relatively large infrastructure costs compared to operating costs. Thus, as it is common with many other 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 demand, a policy rule that is known as Ramsey pricing.

There is a large number of studies focusing on the estimation of the price elasticity of water usage demand; see e.g. Arbués et al. (2003) and Araral and Wang (2013) for excellent surveys. 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. Moreover, 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.

For this reason, in what follows we 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.

5.2. Data and Methodology

We make use of 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). Observations across households are aggregated to some extent in order to maintain privacy. In particular, each cross-sectional unit represents an average of four to six households that

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

are located nearby and have similar property size.[9]

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 contains $4,500$ 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 of the price elasticity 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. Furthermore, the yearly frequency of the data facilitates comparisons with international studies.

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 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.

For the same reason, our sample contains households that have not participated in water appliance efficiency programs; one can anticipate a smaller price elasticity of

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

demand for households that have already participated in such programs for at least two reasons: firstly, these households might be operating under a "conservation mindset" at first place; secondly, they might exhibit a reduced ability to lower their demand due to higher levels of appliance efficiency (demand hardening).

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

$$\ln\left(cons_{i,t}\right) = \alpha \ln\left(cons_{i,t-1}\right) + \beta_1 price_{i,t} + \beta_2 temp_{i,t} + \beta_3 rain_{i,t} + u_{i,t},\ u_{it} = \boldsymbol{\lambda}_i' \boldsymbol{f}_t + \varepsilon_{i,t}, \tag{14}$$

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. This reflects the fact 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 plausible 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 A.1 here]

Figure A.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.

[FIGURE A.1 here]

We specify a log-linear functional form for the following two reasons: contrary to 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 (14) by fitting models with $L = \{0, 1, 2\}$ factors. In addition, we also estimated (14) (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 (13). 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 procedure described in Section 2.

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 A.2 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 model. 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 A.2 here]

Table A.3 presents the estimation results for the coefficients of the model. The last row corresponds to the long-run price coefficient, whose standard error has been obtained using the Delta method.[10]

[TABLE A.3 here]

The sign of the estimated price coefficient is plausible in all cases.[11] 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.

Remark 7. The result of the overidentifying restrictions test statistic for $M1_c$ suggests 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 Online Supplementary Appendix. The 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 is motivated by the fact that the coefficient of price is not statistically significant. The resulting test statistic equals 2.11 and has a p-value close to zero. Hence, there is sufficient statistical evidence to 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 A.4 we present some estimates of the

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

[11] It is worth noting here that the estimates of the price coefficient obtained from MTW_{DIF} and MTW_{SYS} are positive and statistically insignificant in both cases. This demonstrates the importance of allowing for a genuine factor component in the model.

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 A.4 here]

As expected, urban water demand appears to be rather more elastic in the long-run than it is in the short run, which may be attributed to habit formation and technological constraints of water appliance efficiency. 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 within the sample. 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 considerably estimation of dynamic panel data models with multi-factor residuals and fixed T. 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. The simulation study shows that the performance of the proposed estimators is more than satisfactory 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

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. 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.

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.

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 (2015): "Simplified Estimators for Dynamic Panels with a Multifactor Error Structure," Mimeo.

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.

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.

Newey, W. K. and D. McFadden (1994): "Large Sample Estimation and Hypothesis Testing," in *Handbook of Econometrics*, ed. by J. Heckman and E. Leamer, Elsevier, vol. 4, chap. 36, 2111–2245.

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.

Sarafidis, V. and T. J. Wansbeek (2012): "Cross-sectional Dependence in Panel Data Analysis," *Econometric Reviews*, 31, 483–531.

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. Tables and Figures

Figure A.1: Water Consumption, Price and Temperature

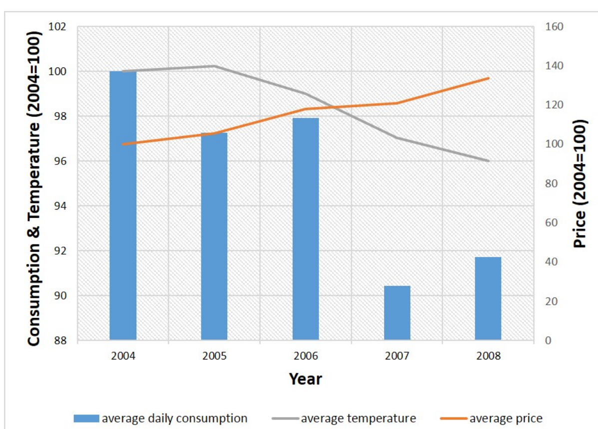

Table A.1: Descriptive Statistics

		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

Table A.2: 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 A.3: Estimation Results for $M1_c$.

	Coef.	**Std. Err.**	**t-ratio**	**p-value**
$\hat{\alpha}$	.393	.056	7.06	.000
$\hat{\beta}_1$	-.213	.072	-2.98	.003
$\hat{\beta}_2$	-.006	.009	-0.73	.466
$\hat{\beta}_3$	.040	.010	3.55	.000
$\frac{\hat{\beta}_1}{1-\hat{\alpha}}$	-.352	.140	-2.50	.012

Table A.4: 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

Appendix B. Monte Carlo results

Table B1: Nonlinear GMM estimator using $w_i = y_{i,0}^0$. $\rho = 0.6$.

Designs					GMM 1 step								GMM 2 step								
					α				β				α				β				J
N	T	α	μ_λ	δ	Bias	RMSE	qStd	Size	Bias	RMSE	qStd	Size	Bias	RMSE	qStd	Size	Bias	RMSE	qStd	Size	Size
200	4	.4	1	.0	-.003	.024	.072	.066	.002	.017	.049	.056	.000	.020	.058	.056	.001	.014	.043	.061	.051
200	4	.4	1	.3	-.008	.038	.113	.061	.004	.026	.076	.062	-.001	.027	.080	.061	.002	.018	.056	.066	.049
200	4	.8	1	.0	-.003	.026	.082	.066	.001	.008	.025	.068	-.001	.021	.063	.070	.001	.007	.022	.062	.045
200	4	.8	1	.3	-.004	.032	.099	.067	.001	.011	.032	.064	-.002	.024	.071	.068	.001	.008	.026	.065	.048
200	8	.4	1	.0	-.002	.015	.045	.058	.002	.012	.038	.061	.000	.012	.036	.103	.001	.010	.032	.117	.035
200	8	.4	1	.3	-.011	.033	.105	.069	.009	.030	.096	.062	-.004	.019	.059	.129	.004	.018	.055	.135	.034
200	8	.8	1	.0	-.002	.014	.042	.052	.001	.007	.021	.058	-.001	.012	.035	.098	.001	.006	.019	.111	.033
200	8	.8	1	.3	-.005	.021	.065	.066	.003	.013	.037	.060	-.002	.015	.046	.117	.001	.010	.029	.121	.035
800	4	.4	1	.0	-.001	.011	.036	.060	.001	.008	.024	.061	.000	.009	.028	.054	.000	.007	.021	.062	.051
800	4	.4	1	.3	-.003	.018	.057	.054	.002	.012	.038	.062	.000	.012	.038	.054	.000	.009	.027	.058	.051
800	4	.8	1	.0	-.001	.014	.040	.054	.000	.004	.013	.055	.000	.010	.029	.048	.000	.004	.011	.052	.052
800	4	.8	1	.3	-.002	.016	.050	.057	.000	.005	.016	.059	.000	.011	.032	.052	.000	.004	.012	.048	.052
800	8	.4	1	.0	.000	.007	.022	.049	.001	.006	.018	.059	.000	.005	.016	.056	.000	.005	.014	.071	.052
800	8	.4	1	.3	-.003	.016	.049	.056	.002	.013	.042	.058	.000	.008	.024	.070	.000	.007	.022	.076	.054
800	8	.8	1	.0	.000	.007	.021	.052	.000	.003	.010	.063	.000	.005	.016	.054	.000	.003	.009	.072	.049
800	8	.8	1	.3	-.001	.011	.033	.052	.000	.006	.018	.051	.000	.006	.019	.065	.000	.004	.012	.069	.047
200	4	.4	0	.0	-.098	.108	.217	.281	.085	.092	.185	.250	-.039	.050	.169	.338	.026	.036	.131	.290	.519
200	4	.4	0	.3	-.212	.227	.399	.364	.153	.170	.346	.320	-.083	.093	.287	.425	.053	.064	.221	.367	.564
200	4	.8	0	.0	-.069	.078	.199	.174	.022	.025	.066	.129	-.035	.048	.170	.292	.009	.015	.053	.206	.513
200	4	.8	0	.3	-.101	.112	.294	.191	.030	.036	.108	.139	-.050	.061	.235	.327	.013	.019	.074	.227	.530
200	8	.4	0	.0	-.129	.132	.185	.466	.173	.175	.254	.518	-.066	.067	.133	.568	.084	.084	.160	.645	.555
200	8	.4	0	.3	-.441	.443	.461	.711	.457	.461	.521	.703	-.284	.284	.332	.787	.276	.277	.352	.792	.683
200	8	.8	0	.0	-.082	.083	.142	.354	.053	.054	.095	.368	-.049	.050	.112	.490	.026	.026	.062	.492	.578
200	8	.8	0	.3	-.220	.221	.334	.511	.138	.139	.274	.489	-.137	.137	.251	.641	.073	.075	.173	.621	.643
800	4	.4	0	.0	-.103	.116	.211	.308	.093	.101	.178	.293	-.024	.039	.155	.403	.017	.028	.121	.358	.708
800	4	.4	0	.3	-.228	.244	.398	.382	.161	.178	.356	.345	-.071	.084	.276	.471	.041	.055	.219	.420	.734
800	4	.8	0	.0	-.067	.077	.172	.224	.024	.026	.058	.197	-.024	.035	.152	.377	.006	.010	.044	.286	.667
800	4	.8	0	.3	-.098	.110	.277	.235	.030	.035	.104	.196	-.033	.048	.213	.395	.008	.013	.065	.305	.686
800	8	.4	0	.0	-.133	.137	.177	.507	.185	.187	.244	.563	-.049	.051	.111	.576	.062	.063	.129	.615	.878
800	8	.4	0	.3	-.468	.469	.464	.753	.487	.489	.524	.744	-.251	.251	.306	.792	.241	.241	.313	.793	.920
800	8	.8	0	.0	-.082	.083	.132	.404	.055	.055	.097	.442	-.037	.037	.096	.506	.018	.018	.050	.519	.851
800	8	.8	0	.3	-.224	.226	.338	.544	.149	.153	.281	.537	-.113	.113	.225	.637	.059	.060	.149	.620	.875

Table B2: Nonlinear GMM estimator using $w_i = y^1_{i,0}$. $\rho = 0.6$.

Designs					GMM 1 step								GMM 2 step								
					α				β				α				β				J
N	T	α	μ_λ	δ	Bias	RMSE	qStd	Size	Bias	RMSE	qStd	Size	Bias	RMSE	qStd	Size	Bias	RMSE	qStd	Size	Size
200	4	.4	1	.0	-.002	.026	.082	.057	.003	.021	.066	.052	-.002	.023	.070	.056	.002	.019	.058	.054	.047
200	4	.4	1	.3	-.007	.044	.135	.064	.005	.032	.103	.052	-.002	.033	.102	.057	.003	.026	.078	.055	.045
200	4	.8	1	.0	-.002	.032	.099	.072	.001	.011	.033	.060	-.002	.024	.077	.068	.001	.010	.029	.065	.045
200	4	.8	1	.3	-.004	.039	.123	.072	.001	.013	.041	.064	-.003	.028	.089	.063	.001	.011	.034	.064	.042
200	8	.4	1	.0	-.002	.015	.046	.057	.003	.015	.043	.061	-.001	.012	.037	.082	.001	.012	.034	.096	.037
200	8	.4	1	.3	-.013	.040	.118	.069	.013	.037	.111	.071	-.005	.020	.060	.111	.005	.019	.057	.113	.032
200	8	.8	1	.0	-.002	.014	.042	.052	.001	.007	.022	.056	-.001	.012	.035	.084	.001	.007	.020	.094	.033
200	8	.8	1	.3	-.005	.022	.068	.063	.003	.014	.040	.058	-.003	.016	.045	.087	.002	.010	.030	.098	.034
800	4	.4	1	.0	.000	.013	.040	.050	.000	.010	.033	.048	.000	.011	.034	.053	.000	.009	.028	.048	.048
800	4	.4	1	.3	-.001	.021	.067	.048	.002	.016	.051	.050	.000	.016	.049	.055	.001	.012	.037	.050	.048
800	4	.8	1	.0	.000	.016	.050	.058	.000	.005	.017	.049	-.001	.011	.035	.052	.000	.004	.014	.050	.043
800	4	.8	1	.3	-.001	.020	.062	.063	.000	.007	.021	.056	.000	.013	.041	.054	.000	.005	.016	.050	.046
800	8	.4	1	.0	-.001	.008	.022	.047	.001	.007	.021	.063	.000	.006	.017	.053	.000	.005	.016	.070	.051
800	8	.4	1	.3	-.005	.019	.058	.056	.004	.018	.055	.061	-.001	.008	.025	.068	.001	.008	.025	.075	.049
800	8	.8	1	.0	-.001	.007	.021	.054	.000	.004	.011	.064	.000	.005	.016	.052	.000	.003	.009	.065	.046
800	8	.8	1	.3	-.001	.011	.033	.050	.001	.007	.020	.054	-.001	.007	.020	.060	.000	.005	.013	.067	.044
200	4	.4	0	.0	-.003	.026	.081	.065	.002	.021	.065	.053	-.001	.024	.071	.063	.001	.020	.060	.059	.044
200	4	.4	0	.3	-.006	.042	.132	.064	.005	.031	.100	.050	-.002	.034	.101	.059	.002	.026	.079	.061	.044
200	4	.8	0	.0	-.004	.032	.100	.074	.001	.011	.035	.058	-.003	.025	.076	.065	.001	.010	.030	.069	.042
200	4	.8	0	.3	-.006	.041	.126	.075	.002	.014	.044	.069	-.004	.028	.089	.069	.001	.012	.035	.064	.046
200	8	.4	0	.0	-.003	.015	.045	.059	.003	.014	.041	.061	-.001	.012	.037	.090	.001	.012	.034	.092	.035
200	8	.4	0	.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	.041	.059	.001	.007	.022	.057	-.001	.012	.036	.084	.001	.007	.021	.090	.036
200	8	.8	0	.3	-.005	.021	.065	.060	.003	.013	.039	.063	-.002	.015	.045	.090	.002	.010	.029	.086	.036
800	4	.4	0	.0	.000	.013	.039	.053	.000	.010	.033	.051	.000	.011	.034	.054	.000	.009	.029	.051	.048
800	4	.4	0	.3	-.002	.020	.064	.046	.001	.016	.049	.051	-.001	.015	.047	.055	.000	.012	.038	.050	.046
800	4	.8	0	.0	-.002	.016	.050	.055	.000	.005	.017	.051	-.001	.012	.036	.051	.000	.005	.014	.055	.046
800	4	.8	0	.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	.022	.049	.001	.007	.020	.062	.000	.006	.017	.054	.000	.005	.016	.070	.052
800	8	.4	0	.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	.021	.054	.000	.004	.011	.057	.000	.006	.016	.054	.000	.003	.009	.066	.048
800	8	.8	0	.3	-.002	.011	.032	.053	.001	.006	.019	.054	-.001	.007	.020	.061	.000	.005	.014	.070	.046