Aggregation of Linear Models for Panel Data

Alexandre Petkovic
ECARES, Université Libre de Bruxelles

David Veredas
ECARES, Université Libre de Bruxelles

ECARES working paper 2009-012
AGGREGATION OF LINEAR MODELS FOR PANEL DATA

Alexandre Petkovic\textsuperscript{1} and David Veredas\textsuperscript{2}

March 2009

Abstract

We study the impact of individual and temporal aggregation in linear static and dynamic models for panel data in terms of model specification and efficiency of the estimated parameters. Model wise we find that i) individual aggregation does not affect the model structure but temporal aggregation may introduce residual autocorrelation, and ii) individual aggregation entails heteroskedasticity while temporal aggregation does not. Estimation wise we find that i) in the static model, estimation by least squares with the aggregated data entails a decrease in the efficiency of the estimated parameters but we cannot rank different aggregation schemes in terms of efficiency, and ii) in the dynamic model, estimation by GMM does not necessarily entail a decrease in the efficiency of the estimated parameters under individual aggregation and no analytic comparison can be established for temporal aggregation, though simulations suggests that temporal aggregation deteriorates the accuracy of the estimates.

Keywords: panel data, temporal aggregation, model specification, efficiency.

JEL classification: C23, C51, C52.

\textsuperscript{1}ECARES, Université Libre de Bruxelles; email: apetkovic@ulb.ac.be.
\textsuperscript{2}ECARES, Université Libre de Bruxelles; email: dveredas@ulb.ac.be.

Corresponding address: David Veredas, ECARES, Université Libre de Bruxelles. 50 Av F.D. Roosevelt CP 114, B1050 Brussels, Belgium. Phone: +3226504218. Fax: +3126504475.

We are grateful to participants of the Third Brussels-Waseda seminar in Time Series and Financial Statistics and to Christian Hafner, Cheng Hsiao, Marc Hallin, Helmut Lütkepohl, Paolo Paruolo and Masanobu Taniguchi for useful comments. We are very grateful to Davy Paindaveine for his detailed reading of the article and sharp comments. Any remaining errors and inaccuracies are ours.
1 Introduction

It is often the case that the econometrician holds a database consisting of a panel in which, for the variable of interest, individuals are aggregated in groups and time frequency is low. We will denote to this panel *aggregated data*. However, the econometrician’s interest may be at individual level and at a higher frequency. We will denote to this panel *disaggregated data*. The econometrician may be tempted to estimate a linear model for panel data, either static or dynamic, for the disaggregated data with the aggregated data, and draw conclusions at disaggregate level. In this article we study the econometric consequences of such a procedure in terms of i) model specification: given a model for the disaggregated data, how does data aggregation affect the model structure?, and ii) parameter estimation: how does data aggregation affect the estimates of the model for the disaggregated data?

The first articles on temporal aggregation track back to the 70’s. Zellner and Montmarquette (1971) study the impact of temporal aggregation on the estimation procedure of the parameters in a linear regression model. Sims (1971) studies the consequences of discretization on the interpretation of coefficients of a linear regression model. The first paper to address temporal aggregation in time series models is Amemiya and Wu (1972). They show that, if the original variable is generated by an AR model of order \(p\), the aggregated variable follows an AR model of order \(p\) with MA residuals structure. Brewer (1973) presents a generalization of the results obtained by Amemiya and Wu for ARMA models with exogenous regressors. Wei (1978) derives the model structure for temporally aggregated data when the high frequency model includes seasonal polynomials. He shows that, if the aggregation frequency is the same as the seasonal frequency (for instance, intra-annual seasonality and annual aggregation), the aggregated model reduces to a model without seasonality. Tiao (1972) investigates temporal aggregation for MA models for integrated time series. Weiss (1984) discusses flow and stock aggregation schemes for ARIMA models. Drost and Nijman (1993) derive the order conditions for temporally aggregated univariate GARCH models. More recently, Jorda and Marcellino (2002) generalized the results of Brewer (1973) by considering temporal aggregation over intervals of random length. Silvetrini and Veredas (2008) survey temporal aggregation in univariate and multivariate ARIMA-GARCH time series models.

The literature of individual aggregation, though it tracks back to earlier dates, is less extensive. Theil (1954) and Zellner (1962) study the effect of individual aggregation in the linear regression models. Chambers (1973) studies how individual aggregation can be responsible for the presence of the long run memory observed in aggregated data. Granger (1980) and Granger (1987) study the impact of individual aggregation in ARMA models. He concludes that, when the individuals have identical models (orders and parameters), the model for the aggregated data is the same. However, as individual heterogeneity increases, the orders increase, long-memory appears, and the model becomes untractable.

Little, if nothing, has been investigated on the aggregation of linear models for panel data. Since panel data are typically indexed by individuals and time, aggregation can take three different (and non exclusive) forms: aggregation over individuals and/or over time. Individual and time aggregation are conceptually different. The key difference is that time conforms an ordered sequence, that is, the time index is not permutation-free. Temporal aggregation may take different schemes: stock (observations are sampled every \(m\) periods), flow (observations are summed every \(m\) periods), average, weighted average, etc. In this article we consider a general aggregation scheme that nests them. However, when necessary, we will exemplify to the above-mentioned special cases. Individual aggregation, by contrast, is more difficult to define as individuals can be permuted without affecting the properties of the panel. We
take the approach of Forni and Lippi (1997) and consider that individuals are aggregated in a pre-defined number of groups.

Starting from a random/fixed and static/dynamic linear model, we study the consequences of using the aggregated data in a model designed for the disaggregated data in terms of i) model specification and ii) parameter estimation.\(^1\)

In terms of model specification, we find that while temporal aggregation does not have any influence in the static case, it generally affects the structure of the model in the dynamic setting, similarly to the findings in ARMA models. Temporal aggregation generally entails MA components. So, for instance, if the disaggregated model is an AR(1) with exogenous regressors, temporal aggregation, in general, introduces residual autocorrelation of order one, as in Amemiya and Wu (1972), and creates a constrained dynamic response with respect to the exogenous variables.\(^2\) These findings hold regardless of whether the model has fixed or random effects. Individual aggregation, by contrast, typically does not entail changes in model specification, but it may create heteroskedasticity.

Estimation wise, aggregation affects the properties of the estimates.\(^3\) In linear models for panel data, parameters are estimated by, mainly, two methods: least squares and GMM. We show that if parameters are estimated by least squares -this is the method we opt for in the static model-, the information loss incurred by aggregation affects the efficiency of the estimates. Since the estimates of the aggregated model are a function of the aggregated data, which, in turn, are an affine transformation of the disaggregated data, we explicit compute the efficiency loss and show that as aggregation increases (both individual and temporal), the quality of the estimated parameters deteriorates. However, we show that theoretically it is not possible to rank aggregation schemes in terms of efficiency loss. Yet, in the simulation study we show that, in general, flow aggregation is less inefficient than stock aggregation. If parameters are estimated by GMM (Arellano-Bond) -this is the method we opt for in the dynamic model-, results are ambiguous. In general, an explicit comparison is not possible but some results can be worked out for individual and temporal aggregation separately. Individual aggregation does not necessarily increase the variance of the estimator. In fact, in some circumstances it may actually improve the estimates. On the other hand, temporal aggregation modifies the number and the form of moment conditions used in GMM (even some moment conditions are nonzero), rendering the estimation problem non-linear and the analytic comparison between the disaggregated and aggregated variances is unfeasible. However numerical simulations suggests that temporal aggregation increases the variance of the estimators.

The paper is organized as follows: Section 2 introduced notation and definitions, Section 3 studies the effect of aggregation on model specification, Section 4 introduces further notation and analyzes the effect of aggregation on parameter estimation. Section 5 concludes. Proofs and preparatory Lemmas are relegated to the Appendix.

---

\(^1\)By dynamic linear model we understand an AR(1) with exogenous variables, which is the most common dynamic model in the analysis of the panels. Other models (i.e. AR of higher order or ARMA) are possible but they are not commonly found in the panel literature.

\(^2\)This result can be seen as a generalization of those obtained in Tilak (1998) and Tilak (2000).

\(^3\)We estimate the parameters of the conditional mean. The variances of the error terms, which are always assumed to be constant in the model for disaggregated data, can be estimated from the residuals.
2 Notation and Definitions

Let \( y_{i,t} \in \mathbb{R} \) be the realization of the random variable of interest observed at time \( t = 1, \ldots, N \) for individual \( i = 1, \ldots, I \). Likewise for \( x_{i,t} \in \mathbb{R} \), an exogenous regressor. We assume that the relation between \( y_{i,t} \) with its own past and \( x_{i,t} \) is given by a dynamic linear model with either fixed or random effects:

\[
\begin{align*}
y_{i,t} &= \alpha_i + \gamma y_{i,t-1} + \beta x_{i,t} + u_{i,t}, \quad (1) \\
y_{i,t} &= \mu + \gamma y_{i,t-1} + \beta x_{i,t} + \alpha_i + u_{i,t}, \quad (2)
\end{align*}
\]

where \( u_{i,t} \sim iid(0, \sigma_u^2) \) and \( \gamma \in ]-1, 1[ \), \( \mu \in \mathbb{R} \), \( \beta \in \mathbb{R} \) and \( \sigma_\alpha^2 \in ]0 + \infty[ \). In (1) the individual effects \( \alpha_i \in \mathbb{R} \) are fixed, and hence parameters to be estimated, while in (2) are assumed to be \( iid \) random variables with zero mean and variance \( \sigma_\alpha^2 \in ]0 + \infty[ \). For the sake of simplicity we do not consider richer models that include, for instance, time effects, time-invariant and/or individual-invariant regressors, and more than one exogenous variable. These extensions, although technically feasible, would imply tedious calculations.

Sample information for \( y_{i,t} \) is assumed to be available only every \( m \) periods \((m, 2m, 3m, \ldots)\) -where \( m \), an integer value larger that one, is the aggregation frequency- and in \( A \) groups of individuals.

Individual aggregation is linear and of the form:

\[
\tilde{y}_{a,t} = \left( \sum_{i=1}^{I} M_a^i \right) y_{i,t} = B(a) y_{i,t},
\]

where \( a = 1, \ldots, A \) refers to the group and \( M_a^i \) is the weight, assumed to be known, associated with individual \( i \), being zero if an individual does not belong to group \( a \). The group \( a \) is composed by \( I_a \) individuals, and the size of the \( A \) groups can be different but known. The use of this aggregation scheme is justified by the fact that individuals are not sequentially ordered when they are aggregated. We assume that an individual cannot belong to different groups, i.e. \( \sum_{i=1}^{I} M_a^i M_{a'}^i = 0 \ \forall a' \neq a \). This can be however relaxed. The consequences of relaxing this assumption are on model specification but not on estimation since it would entail correlation between groups that can be handled by the generalized least squares or GMM estimators. This scheme embeds some important aggregation schemes: i) Flow: \( M_a^i = 1 \) or 0 with \( \sum_{a=1}^{A} \sum_{i=1}^{I} M_a^i = I \). That is, sum of individuals belonging to group \( a \). ii) Stock: \( M_a^i = 1 \) or 0 and \( \sum_{i=1}^{I} M_a^i = 1 \). That is, groups with only one individual that are selected according to some pre-specified criteria. iii) Per capita: \( M_a^i = (1/I_a) \) or 0 and \( \sum_{a=1}^{A} \sum_{i=1}^{I} M_a^i = A \). That is, the total sum of individuals is divided by the number of individuals in the group.

Temporal aggregation is also linear and of the form:

\[
\tilde{y}_{i,t} = \left( \sum_{j=0}^{m-1} w_j L^j \right) y_{i,t} = T(L) y_{i,t},
\]

where \( L \) is the lag operator and \( w_j \) is the weight, assumed to be known, attached to the observation at time \( t - j \). This scheme embeds aggregation schemes similar to the ones of individual aggregation: i) Flow: \( w_j = 1 \), for \( j = 0, \ldots, m - 1 \). That is, aggregation of \( y_{i,t} \) is carried out over \( m \) periods. ii) Stock: \( w_0 = 1 \) and \( w_j = 0 \) for \( j = 1, \ldots, m - 1 \). One every \( m \) observations is kept, the rest being skipped. iii) Average: \( w_j = \frac{1}{m} \), for \( j = 0, \ldots, m - 1 \). iv) Weighted average: \( w_j = \chi_j \), for \( j = 0, \ldots, m - 1 \), where \( \chi_j \) are the weights that sum one.
Note that flow, averaging and weighted averaging aggregation schemes are rolling sums. In other words, \( \tilde{y}_{i,t} \) is indexed by \( t \), which means a sequence of sums that overlap over \( m - 1 \) periods. However, the aggregated data does not overlap. To indicate the aggregated data we introduce another time scale, \( T \), that runs in \( mt \) periods. So that \( t = 1, 2, \ldots, N \), while \( T = m, 2m, \ldots, m\tau \). Thus, \( \tau = N/m \) and we subindex the temporally aggregated series by \( T \): \( \tilde{y}_{i,T} = \tilde{y}_{i,mt} \).

Combining these two aggregation schemes, we compute the aggregated data as:

\[
\tilde{y}_{a,T} = T(L)B(a)y_{i,t},
\]

and likewise for \( x_{i,t} \) and \( u_{i,t} \). We denote \( T(L)B(a) \) as the General Static Aggregation Scheme (henceforth GSAS) and we use in the static model \( (\gamma = 0) \). The term General stands for simultaneous temporal and individual aggregation.

For the dynamic model, since the dependent variable is present on both sides of the model, we first express (1) and (2) in terms of the lag operator:

\[
\begin{align*}
(1 - \gamma L) y_{i,t} &= \alpha_i + \beta x_{i,t} + u_{i,t}, \\
(1 - \gamma L) y_{i,t} &= \mu + \beta x_{i,t} + \alpha_i + u_{i,t}.
\end{align*}
\]

And we apply the aggregation scheme \( (1 - \gamma^m L^m)(1 - \gamma L)^{-1}T(L)B(a) \), which we denote as the General Dynamic Aggregation Scheme (henceforth GDAS) and it follows Brewer (1973). So, for instance, the aggregated left-hand side of the model is

\[
\left( \frac{1 - \gamma^m L^m}{1 - \gamma L} \right) T(L)B(a)(1 - \gamma L) y_{i,t} = (1 - \gamma^m L^m)\tilde{y}_{a,T}.
\]

The first term of the left-hand side is a ratio of two polynomials. The denominator contains the inverted roots of the AR part of the model and its numerator contains the same roots, but powered by the aggregation frequency. This term ensures that the powers of the lag operator \( L \) of the model for aggregated data are only divisible by \( m \), as it can be seen on the right-hand side.

### 3 The effect of aggregation on model specification

In this section we study the effect of not observing \( y_{i,t} \) but \( \tilde{y}_{a,T} \) on model specification. In other words, if we know the model for \( y_{i,t} \), how is specification affected by the fact that we only observe \( \tilde{y}_{a,T} \)? We first study the static case, followed by the dynamic model.

#### 3.1 Static Model

We consider a static fixed effects model, i.e. (1) with \( \gamma = 0 \):

\[
y_{i,t} = \alpha_i + \beta x_{i,t} + u_{i,t}.
\]

Applying GSAS on both sides of the model:

\[
\tilde{y}_{a,T} = \left( \sum_{j=0}^{m-1} w_j \right) \alpha_a + \tilde{x}_{a,T} \beta + \tilde{u}_{a,T},
\]
where \( \alpha_a = \left( \sum_{i=1}^I M_i^a \alpha_i \right) \), \( \tilde{u}_{a,T} \) is independent with \( E(\tilde{u}_{a,T}) = 0 \) and

\[
Var(\tilde{u}_{a,T}) = \sigma_u^2 \left( \sum_{i=1}^I (M_i^a)^2 \right) \left( \sum_{j=0}^{m-1} w_j^2 \right).
\]

This result leads to the following conclusions. First, the slope parameter \( \beta \) is innocuous to temporal aggregation. This result is not surprising since we assume that the slopes are constant through time and individuals. Second, the fixed effects are affected by aggregation in three ways: i) the number of fixed effects reduces to the number of aggregated groups, from \( I \) to \( A \), ii) temporal aggregation entails a re-scaling of the same magnitude to all the effects, iii) \( \alpha_a \) is a weighted average of the disaggregated effects. Third, the aggregated error is not \( iid \) as it shows heteroskedasticity, unless \( \sum_{i=1}^I (M_i^a)^2 = \sum_{i=1}^I (M_i^{a'})^2 \), \( \forall a,a' \). This may happen in different scenarios. The easiest one is when individuals are aggregated in groups of equal size \( (I_a = I_{a'}) \) and the weights are homogenous across groups. The heteroskedasticity increases with more heterogenous group sizes, more heterogenous weights across groups or both.

We now consider a static random effects model, i.e. (2) with \( \gamma = 0 \):

\[
y_{i,t} = \mu + \beta x_{i,t} + \alpha_i + u_{i,t}.
\]

Applying GSAS on both sides of the model:

\[
\tilde{y}_{a,T} = \left( \sum_{j=0}^{m-1} \gamma^j \right) \left( \sum_{j=0}^{m-1} w_j \right) \mu_a + \tilde{x}_{a,T} \beta + \tilde{a}_a + \tilde{u}_{a,T},
\]

where \( \mu_a = \left( \sum_{i=1}^I M_i^a \right) \mu, \tilde{a}_a, \tilde{u}_{a,T} \) is the same as in the fixed effects model, \( E(\tilde{a}_a) = 0 \), and

\[
Var(\tilde{a}_a) = \sigma_a^2 \left( \sum_{i=1}^I (M_i^a)^2 \right) \left( \sum_{j=0}^{m-1} w_j \right)^2.
\]

As in the fixed effects model, the slope parameter \( \beta \) is innocuous to aggregation and individual aggregation causes heteroskedasticity in the aggregated error terms \( \tilde{a}_a \) and \( \tilde{u}_{a,T} \) unless \( \sum_{i=1}^I (M_i^a)^2 = \sum_{i=1}^I (M_i^{a'})^2 \), \( \forall a,a' \). The aggregated intercept is group-specific unless \( \sum_{i=1}^I M_i^2 = \sum_{i=1}^I M_i^{a'} \), \( \forall a,a' \). That is, by aggregating individuals in a random effects model, there appears both random and fixed effects.

### 3.2 Dynamic Model

We consider the dynamic fixed effects model (3) and the GDAS. Similarly to the static case, we premultiply the model by the aggregation scheme yielding:

\[
\tilde{y}_{a,T} = \left( \sum_{j=0}^{m-1} \gamma^j \right) \left( \sum_{j=0}^{m-1} w_j \right) \alpha_a + \gamma^m \tilde{y}_{a,T-m} + \tilde{x}_{a,T-m+1} \beta + \ldots + \tilde{x}_{a,T-m+1} \beta \gamma^{m-1} + \eta_{a,T},
\]

where \( \alpha_a = \left( \sum_{i=1}^I M_i^a \alpha_i \right) \) and \( \eta_{a,T} = \tilde{u}_{a,T} + \gamma \tilde{u}_{a,T-1} + \ldots + \gamma^{m-1} \tilde{u}_{a,T-m+1} \) is a zero mean error term with variance

\[
\sigma_\eta^2 = \sigma_u^2 \left( \sum_{i=1}^I (M_i^a)^2 \right) \left[ w_0^2 + \sum_{p=1}^{m-1} \left( \sum_{j=0}^{m-1+p} w_j \gamma^{m-1+p-j} \right)^2 \right].
\]
This result leads to the following conclusions. First, the fixed effects are affected by aggregation in the same way as in the static model (the term on $\gamma$ has the same consequences as the term on $w_j$). Second, temporal aggregation creates a dependence with respect to past aggregated exogenous variables. However the dependence is constrained: the coefficients attached to the lagged aggregated exogenous variables are nonlinear functions of $\beta$ and $\gamma$ (similarly to Palm and Nijman, 1984). Only if the temporal aggregation scheme is stock (where $w_0 = 1$ and $w_j = 0 \forall j > 0$) the dependence with lagged aggregated exogenous variables disappears. Third, similarly to Amemiya and Wu (1972), temporal aggregation entails autocorrelation of order one in $\eta_{a,T}$:

\[
E(\eta_{a,T}\eta_{a,T-m}) = \sigma_a^2 \left( \sum_{i=1}^{l} (M_{i}^a)^2 \right) \left[ \sum_{v=0}^{m-2} \left( \sum_{j=0}^{v} w_j \gamma^{v-j} \right) \left( \sum_{j=v+1}^{m-1} w_j \gamma^{m+v-j} \right) \right] \\
E(\eta_{a,T}\eta_{a,T-sm}) = 0 \quad s = \pm 2, \pm 3, \pm 4, ...
\]

In other words, the dynamic model specification is affected by temporal aggregation: from an AR(1) to an ARMA(1,1), except if aggregation is stock. Last, and as in the static case, the aggregated error shows heteroskedasticity unless $\sum_{i=1}^{l} (M_{i}^a)^2 = \sum_{i=1}^{l} (M_{i}^{a'})^2$, $\forall a, a'$.

Finally, we consider the dynamic random effects model (4) and the GDAS. Premultiplying the model by the aggregation scheme yields:

\[
\tilde{y}_{a,T} = \left( \sum_{j=0}^{m-1} \gamma^j \right) \left( \sum_{j=0}^{m-1} w_j \right) \mu_a + \gamma^m \tilde{y}_{a,T-m} + \tilde{x}_{a,T} \beta + \ldots + \tilde{x}_{a,T-m+1} \beta \gamma^{m-1} + \tilde{\alpha}_a + \eta_{a,T}
\]

where $\mu_a = \left( \sum_{i=1}^{l} M_i^a \right) \mu$, $\tilde{\eta}_{a,T}$ is the same as in the fixed effects model, $E(\tilde{\alpha}_a) = 0$, and

\[
Var(\tilde{\alpha}_a) = \sigma_a^2 \left( \sum_{i=1}^{l} (M_{i}^a)^2 \right) \left( \sum_{j=0}^{m-1} w_j \right)^2 \left( \sum_{j=0}^{m-1} \gamma^j \right)^2.
\]

Conclusions from applying GDAS to (2) are a combination of those from the random effects static model the fixed effects dynamic model: dependence of $\tilde{y}_{a,T}$ with respect to past aggregated exogenous variables, heteroskedastic and serially correlated aggregated errors $\eta_{a,T}$, and group-specific intercepts capturing group heterogeneity. The aggregated model boils down to the disaggregated model if temporal and individual aggregation are stock.

### 4 The effect of aggregation on estimation

In this section we study the consequences of estimating the parameters, namely $\mu$, $\beta$ and $\gamma$, of the models for $y_{i,t}$ but with $y_{a,T}$. We don’t focus on the estimation of the fixed effects as they are easily estimated once the slope parameters are estimated. We first introduce further notation, rewriting (1)-(4), GSAS and GDAS in matrix notation. In the second part of the section we focus on estimation of $\mu$ and $\beta$ in the static model by least squares and we show how to explicitly compute the efficiency loss. We also show that it is not possible to compute the efficiency losses entailed by two different aggregation schemes.

In the third part we study estimation of $\beta$ and $\gamma$ in the dynamic model by GMM, namely Arellano-Bond. Because individual and temporal aggregation have different consequences on
the model specification, we study them separately. In the previous section we saw that individual aggregation only affects the scale of the model leaving its structure unchanged. This means that moments conditions of the GMM remains valid for the aggregated data. However, temporal aggregation changes the model specification (it generates a dynamic response to the means that moments conditions of the GMM remains valid for the aggregated data. However, individual aggregation only affects the scale of the model leaving its structure unchanged. This the model specification, we study them separately. In the previous section we saw that individual aggregation only affects the scale of the model leaving its structure unchanged. This means that moments conditions of the GMM remains valid for the aggregated data. However, temporal aggregation changes the model specification (it generates a dynamic response to the exogenous variables and may introduce autocorrelation in the aggregated residuals) which may imply a modification of the moments conditions.

4.1 Notation

The fixed effects model (1) can be written as

\[
\begin{pmatrix}
  y_1 \\
  y_2 \\
  \vdots \\
  y_I
\end{pmatrix}
= \begin{pmatrix}
  \alpha_1 e_N \\
  \alpha_3 e_N \\
  \vdots \\
  \alpha_I e_N
\end{pmatrix} + \gamma
\begin{pmatrix}
  y_{1,-1} \\
  y_{2,-1} \\
  \vdots \\
  y_{I,-1}
\end{pmatrix} + \beta
\begin{pmatrix}
  x_1 \\
  x_2 \\
  \vdots \\
  x_I
\end{pmatrix} +
\begin{pmatrix}
  u_1 \\
  u_2 \\
  \vdots \\
  u_I
\end{pmatrix},
\]

where \( y_i \) is a \( N \times 1 \) vector with the individual observations of individual \( i \), \( \alpha_i \) is the fixed effect of individual \( i \), \( e_N \) is a \( N \times 1 \) vector of ones, \( y_{i,-1} = (y_{i,0}, \ldots, y_{i,N-1})' \) is a \( N \times 1 \) vector of lagged \( y_{i,t} \), \( x_i \) is a \( N \times 1 \) vector containing the individual regressor and \( u_i \) is a \( N \times 1 \) vector with the iid individual error terms with \( E(u_i) = 0 \) and \( E(u_i u_i') = \sigma_u^2 I_N \), where \( I_N \) is an identity matrix of size \( N \). The above equation can be written more compactly as

\[
Y = \alpha + \gamma Y_{-1} + X \beta + U,
\]

where \( U \) is a \( I N \times 1 \) vector of iid errors such that \( E(U) = 0 \) and \( E(UU') = \sigma_u^2 I_N \).

Likewise, the random effects model (2) can be written as

\[
Y = Z \delta + W,
\]

where \( Z = (e_{IN}, Y_{-1}, X) \), \( \delta = (\mu, \gamma, \beta)' \), \( e_{IN} \) is a \( I N \times 1 \) vector of ones, and

\[
W = \begin{pmatrix}
  u_1 + \alpha_1 e_N \\
  u_2 + \alpha_2 e_N \\
  \vdots \\
  u_I + \alpha_I e_N
\end{pmatrix}
\]

where \( \alpha_i \) are iid random variables with zero mean and variance \( \sigma_\alpha^2 \). The variance-covariance matrix of \( u_i + \alpha_i e_N \) is

\[
\Sigma = \sigma_u^2 I_N + \sigma_\alpha^2 e_N e_N'.
\]

Models (3) and (4) can also be expressed in matrix form:

\[
(1 - \gamma L) Y = \alpha + X \beta + U \
\]

and

\[
(1 - \gamma L) Y = Z' \delta + W,
\]

where \( Z' = (e_{IN}, X) \) and \( \delta' = (\mu, \beta)' \).

We now express GSAS in vectorial form. Following Lutkepohl (1986), let \( F_1 \) be a \( \tau \times N \) matrix such that

\[
F_1 = \begin{pmatrix}
  \omega' & 0 & \ldots & 0 \\
  0 & \omega' & \ldots & 0 \\
  \vdots & \vdots & \ddots & \vdots \\
  0 & 0 & \ldots & \omega'
\end{pmatrix}
\]

...
with $\omega' = (w_{m-1}, \ldots, w_0)$. The vector $F_1 y_i$ therefore contains the temporally aggregated observations for individual $i$. Let

$$F_2 = \begin{pmatrix}
M^1_1 & M^1_2 & \cdots & M^1_N \\
M^2_1 & M^2_2 & \cdots & M^2_N \\
\vdots & \vdots & \ddots & \vdots \\
M^A_1 & M^A_2 & \cdots & M^A_N
\end{pmatrix}$$

be a $A \times I$ matrix such that $(F_2)_{vl} = M^v_l$ i.e. the $vl$ entry is equal to the weight of individual $l$ in group $v$. Hence the vector of individually aggregated data can be written as $F_2 \otimes I_N Y$, where $\otimes$ denotes the Kronecker product. The GSAS can be rewritten as $FY = (F_2 \otimes F_1) Y$, where $F$ is a $A \tau \times IN$ matrix, that is a matrix with rows equal to the total number of aggregated observations and columns equal to the total number of disaggregated observations.

The GDAS can be written as an extension of GSAS, by just introducing the ratio of autoregressive polynomials:

$$\left(1 - \frac{(\gamma L)^m}{1 - \gamma L}\right) (F_2 \otimes F_1).$$

Note that while GSAS is applied to the static models (5) and (6), in order to apply GDAS we have to express the models as (7) and (8).

### 4.2 Static Fixed Effects model

Consider (5) with $\gamma = 0$. If we could observe $y_{i,t}$, $\beta$ could be estimated as usual, by first pre-multiplying the model by the projection matrix

$$\tilde{Q}_{N,I} = I_I \otimes Q_N \text{ where}$$

$$Q_N = I_N - \frac{1}{N} e_N e'_N$$

centers the observations with respect to the individual means. The ordinary least squares estimator is

$$\hat{\beta} = (X' \tilde{Q}_N X)^{-1} (X' \tilde{Q}_N Y),$$

with variance

$$Var(\hat{\beta}|X) = \sigma_u^2 (X' \tilde{Q}_N X)^{-1}. \quad (9)$$

But since we don’t observe $y_{i,t}$, but $\tilde{y}_{a,T}$, we multiply (5) by GSAS:

$$FY = F\alpha + FX\beta + FU.$$

And we can, in principle, estimate $\beta$ by ordinary least squares by first pre-multiplying the model by an equivalent projection matrix to the previous one but for aggregated data

$$\tilde{Q}_{\tau,A} = I_A \otimes Q_\tau \text{ where}$$

$$Q_\tau = I_\tau - \frac{1}{\tau} e_\tau e'_\tau,$$

However, this method produces inefficient estimators as

$$E(FUU'F') = \sigma_u^2 (F_2 F_2' \otimes F_1 F_1') = \sigma_u^2 \left( \sum_{j=0}^{m-1} w_j^2 \right) (F_2 F_2' \otimes I_\tau)$$
has different diagonal elements unless \( \sum_{i=1}^{I}(M_i) = \sum_{i=1}^{I}(M_i) = 1 \) (M_a_i). That is, it is the term \( \text{F}_2 \text{F}_2' \otimes \text{I}_r \) that produces heteroskedasticity. But since we know the exact form of heteroskedasticity, it is simple to build a generalized least squares estimator. Multiplying both sides of the model by the matrix \( \text{V} = ((\text{F}_2 \text{F}_2')^{-\frac{1}{2}} \otimes \text{I}_r) \):\(^4\)

\[
\text{V} \text{F} \text{Y} = \text{V} \text{F} \alpha + \text{V} \text{F} \text{X} \beta + \text{V} \text{U},
\]

renders \( \text{V} \text{U} \) homokedastic:

\[
\text{E} (\text{V} \text{U} \text{U}' \text{V}') = \text{V} \text{F} \text{E} (\text{U} \text{U}') \text{F}' \text{V}' = \sigma_u^2 (\text{V} \text{F} \text{F}' \text{V}')
\]

\[
= \sigma_u^2 \left( \left( (\text{F}_2 \text{F}_2')^{-\frac{1}{2}} \text{F}_2 \text{F}_2' (\text{F}_2 \text{F}_2')^{-\frac{1}{2}} \right) \otimes \text{F}_1 \text{F}_1' \right)
\]

\[
= \sigma_u^2 \left( \sum_{j=0}^{m-1} w_j^2 \right) (\text{I}_A \otimes \text{I}_r).
\]

And we estimate by ordinary least squares by applying the projection matrix \( \hat{\text{Q}}_{r,A} \):

\[
\hat{\text{Q}}_{r,A} \text{V} \text{F} \text{Y} = \hat{\text{Q}}_{r,A} \text{V} \text{F} \alpha + \hat{\text{Q}}_{r,A} \text{V} \text{F} \text{X} \beta + \hat{\text{Q}}_{r,A} \text{V} \text{U},
\]

The estimator is

\[
\hat{\beta} = (\text{X}' \text{F}' \text{V} \hat{\text{Q}}_{r,A} \text{V} \text{F} \text{X})^{-1} (\text{X}' \text{F}' \text{V} \hat{\text{Q}}_{r,A} \text{V} \text{F} \text{Y}),
\]

with variance

\[
\text{Var}(\hat{\beta} \mid \text{X}) = \sigma_u^2 \left( \sum_{j=0}^{m-1} w_j^2 \right) (\text{X}' \text{F}' \text{V} \hat{\text{Q}}_{r,A} \text{V} \text{F} X)^{-1}.
\]

Variances (9) and (11) are conditional to \( \text{X} \). From the variance decomposition

\[
\text{Var}(\hat{\beta}) = E \left[ \text{Var}(\hat{\beta} \mid \text{X}) \right] + \text{Var} \left( E(\hat{\beta} \mid \text{X}) \right),
\]

which equals \( E \left[ \text{Var}(\hat{\beta} \mid \text{X}) \right] \) since \( E(\hat{\beta} \mid \text{X}) = \beta \). Likewise for \( \text{Var}(\hat{\beta}^a) \). Thus, to show \( \text{Var}(\hat{\beta}^a) \geq \text{Var}(\hat{\beta}) \) it is sufficient to show that \( \text{Var}(\hat{\beta}^a \mid \text{X}) \geq \text{Var}(\hat{\beta} \mid \text{X}) \) for all \( \text{X} \).

The following Theorem shows that temporal and individual aggregation increase the variance of the estimated coefficient and thus results in an efficiency loss.

**Theorem 1** In the fixed effects model (5) with \( \gamma = 0 \), temporal and individual aggregation results in an efficiency loss of the least squares estimator, i.e. the difference \( \text{Var}(\hat{\beta}^a) - \text{Var}(\hat{\beta}) \) is semi-positive definite.

The choice of the aggregation scheme may be given by the problem at hand, but often it is a subjective choice. It is hence of interest to know if some temporal aggregation schemes are less inefficient than others. For example, does stock aggregation entail a higher efficiency loss than flow aggregation? In the sequel we show that there is no a clear answer to this question. Though we know from Theorem 1 that estimation with aggregated data entails an

\(^4\text{Note that compute } \text{V} \text{ is straightforward as } \text{F}_2 \text{F}_2' \text{ is a diagonal matrix (individuals can only belong to one group).} \)
efficiency loss in the estimated parameters, we cannot rank aggregation schemes in terms of efficiency loss. We first show the result for temporal aggregation, followed by the counterpart for individual aggregation.

Let $F_1$ and $F_1^*$ be the matrices corresponding to two different temporal aggregation schemes at the same frequency. Then, computations similar to those in the proof of Theorem 1 show that the aggregation scheme $F_1$ is more efficient than aggregation scheme $F_1^*$ if and only if the matrix

$$F_1'Q_\tau F_1 - F_1'^*Q_\tau F_1^*,$$

(12)

is a semi-positive definite where

$$F_1 = \frac{1}{\sqrt{\sum_{j=0}^{m-1} w_j^2}} F_1$$

And likewise for $F_1^*$. The next Proposition shows that this difference is an indefinite matrix.

**Proposition 1** Let the fixed effects model (5) with $\gamma = 0$ and assume two temporal aggregation schemes $F_1$ and $F_1^*$ characterized by the weights $\omega$ and $\omega^*$. If $\omega \neq \lambda \omega^*$ for all $\lambda \in \mathbb{R}\setminus\{0\}$ then (12) is indefinite. And if $\exists \lambda \in \mathbb{R}\setminus\{0\}$ such that $\omega = \lambda \omega^*$, then (12) equals zero.

This proposition basically says that two different aggregation schemes define directions in hyperplanes that are not parallel, and hence they cannot be compared. The following example shows a simple case where this happens.

**Example 1** To fix ideas, consider the following example. Let $m = 2$ and consider the flow and stock aggregation schemes characterized by the sequence of weights $\omega' = (1, 1)$ and $\omega^* = (1, 0)$ respectively. It can be seen that $\omega' = (a, -a)$ with $a \in \mathbb{R}\setminus\{0\}$ since $\omega'\omega^c = a - a = 0$. Thus the condition $\omega\omega^c$ defines a line in a two dimensional plane. Similarly, the condition $\omega'^*\omega^{*c} = 0$ leads to $\omega^{*c} = (0, a)'$ with $a \in \mathbb{R}\setminus\{0\}$, and thus defines another line in the two dimensional plane. Since $\omega \neq \lambda \omega^*$ for all $\lambda \in \mathbb{R}\setminus\{0\}$ the two lines are not parallel.

Proposition 1 is counter intuitive. It is natural to think that flow aggregation is more appropriate than stock aggregation as it sums observations instead of sampling them. However, Proposition 1 shows that when it comes to parameter estimation this does not need to be true. To clarify this point consider the following example.

**Example 2** Consider model (5) with $\gamma = 0$:

$$y_{i,t} = \alpha_i + x_{i,t}\beta + u_{i,t},$$

and let $x_{i,t} = -x_{i,t-1}$ for all $i$. If data are aggregated as flow with $m = 2$, the aggregated model is

$$y_{i,t} + y_{i,t-1} = 2\alpha_i + (x_{i,t} + x_{i,t-1})\beta + u_{i,t} + u_{i,t-1} = 2\alpha_i + (-x_{i,t-1} + x_{i,t-1})\beta + u_{i,t} + u_{i,t-1} = 2\alpha_i + u_{i,t} + u_{i,t-1}.$$

In such situation $\beta$ is not identified. However, we do not have this problem under stock aggregation.
The computation that lead to the proof of Theorem 1 can also be used to compare different individual aggregation schemes. Specifically, let $F_2$ and $F_2^*$ be two matrices representing two different individual aggregation schemes. Here we assume that the total number of aggregated individuals is fixed and equal to $A$. Let $P_2 = F_2'(F_2F_2')^{-1}F_2$ and likewise for $P_2^*$. The aggregation scheme $F_2$ is more efficient than the aggregation scheme $F_2^*$ if the matrix

$$P_2 - P_2^*,$$

is semi-positive definite. The matrix $P_2$ is the projection matrix in the column space of $F_2'$ and $P_2'$ is the projection matrix in the column of $F_2^*$. Unfortunately, the column spaces of $F_2'$ and $F_2^*$ do not necessarily coincide, as shown in the following Proposition.

**Proposition 2** Let $F_2$ and $F_2^*$ be two individual aggregation schemes such that the number of aggregated groups is the same for both schemes. Consider the matrix $A$ obtained by staking the matrices $F_2$ and $F_2^*$

$$A = \begin{pmatrix} F_2 \\ F_2^* \end{pmatrix}.$$ 

If $\text{rank}(A) > A$, then (13) is indefinite.

**Example 3** In order to fix ideas, consider for example the two following individual aggregation schemes:

$$F_2 = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{pmatrix} \quad \text{and} \quad F_2^* = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}.$$ 

In $F_2$ individual are aggregated as flows while in they are aggregated as stock $F_2^*$ and $\text{rank}(A) = 4$. Thus stock aggregation is not necessarily more inefficient than flow aggregation.

As for Proposition 1, it is intuitive that individual flow aggregation should be more efficient than individual stock aggregation. However, in the following example we show why this intuition is not necessarily true.

**Example 4** Consider model (5) with $\gamma = 0$:

$$y_{i,t} = \alpha_i + x_{i,t}\beta + u_{i,t},$$

where $i = 1, 2, 3, 4$. Assume that the 4 individuals are aggregated in two groups. The first group contains individuals 1 and 2, the second group individuals 3 and 4. Assume furthermore that $x_{1,t} = -x_{2,t}$ and $x_{3,t} = -x_{4,t}$. Similarly to Example 2 it can be seen that stock individual aggregation will be more efficient than flow aggregation.

We illustrate the efficiency loss implied by aggregation with a simulation study. It is based on comparing (9) and (11) for different aggregation schemes. We consider 12 individuals and 40 time periods ($I = 12$ and $N = 40$), three temporal aggregation schemes and four individual aggregation schemes. Time is aggregated as flow ($w_j = 1 \forall j$) and as stock ($w_1 = 1$ and $w_j = 0 \forall j > 1$) and the aggregation schemes are i) type 1: $m = 10$ or $\tau = 4$, ii) type 2: $m = 5$ or
\( \tau = 8, \) iii) type 3: \( m = 4 \) or \( \tau = 10, \) and iv) type 4: \( m = 2 \) or \( \tau = 20. \) That is, the aggregation frequency increases with the type. Individuals are aggregated as flows (i.e. \( M^a_i = 1 \) \( \forall a \) and all \( i \in a \)) and the aggregation schemes are i) type 1: \( A = 3, I_a = 4, \forall a, \) ii) type 2: \( A = 4, I_a = 3, \forall a, \) and iii) type 3: \( A = 6, I_a = 2, \forall a. \) That is, we aggregate in groups of equal size and the group size decreases with the type.

The regressors \( x_{i,t} \) are generated according to a AR(1) processes \( x_{i,t} = \phi x_{i,t-1} + \epsilon_{i,t} \) where \( \phi \) takes values 0.8, 0 and -0.8 and \( \epsilon_{i,t} \sim iidN(0,1). \) Thus we assume that the regressors are independent through individuals but dependent through time. We set \( \sigma^2_u = 1. \) We generate 10000 disaggregated samples for the regressors and we compute the sample mean of the variances (9) and (11). Note that to compute these variances we don’t need to simulate \( y_{i,t} \) nor to estimate (hence avoiding the uncertainty due to estimation). To compute (9) and (11) we only need \( \sigma^2_u, \) the aggregation schemes and to simulate the regressors.

Table 1 is divided in three panels, each divided in three sub-panels. The upper panel shows the results for \( \phi = 0.8, \) the middle for \( \phi = -0.8 \) and the bottom for \( \phi = 0. \) For each sub-panel the upper and middle parts show the average relative efficiency loss when temporal aggregation is flow and stock. The bottom part shows the percentage of times that the variances using the aggregated data under the temporal stock scheme are smaller than under the temporal flow scheme. In all cases, aggregation entails an important loss of efficiency, even if aggregation is mild, confirming Theorem 1. On the other hand, results confirm Proposition 1. There is no temporal aggregation scheme dominates another.\(^5\) When \( \phi \) is large and positive (negative respectively) flow (stock respectively) temporal aggregation entails less efficiency loss. In fact, from the bottom parts of the sub-panels we conclude that flow (stock respectively) aggregation tends to be more efficient when \( \phi \) is large and positive (negative respectively). This is related with the example 2 above. If a positive value for \( x_{i,t} \) tends to be followed by another positive value, we loose more information sampling than summing every \( m \) periods. However, if a positive value for \( x_{i,t} \) tends to be followed by a negative value, summing cancels the observations and the information loss is more important than sampling. In the case that the regressor is a white noise, stock and flow aggregation does not make any difference and the proportion of times that one variance is larger than other is around 50%.

### 4.3 Static Random Effects Model

We now consider the static random effects model: (6) with \( \gamma = 0. \) Since the effects are now part of the error term, the variance-covariance of \( U \) is not diagonal. If we could observe \( y_{i,t}, \) \( \delta \) could be estimated by generalized least squares:

\[
\hat{\delta} = \left( Z' V^2_N Z \right)^{-1} \left( Z' V^2_N Y \right),
\]

where the weighting matrix \( V^2_N \) is given by (see e.g. Hsiao, 2003)

\[
V^2_N = I_N \otimes \frac{1}{\sigma^2_u} \left( I_N - \frac{\sigma^2_u}{\sigma^2_u + N\alpha^2} e_N e'_N \right).
\]  \( \text{(14)} \)

And its variance is

\[
Var(\hat{\delta}|Z) = \left( Z' V^2_N Z \right)^{-1}.
\]  \( \text{(15)} \)

Since we don’t observe \( y_{i,t}, \) but \( y_{a,T}, \) we multiply (5) by GSAS:

\[
FY = FZ\delta + FU.
\]

\(^5\) For the ease of exposition, all over the simulation studies for the static models we consider flow individual aggregation. Results for stock individual aggregation confirm Proposition 2. They are available under request.
Table 1: Efficiency loss - Model (5) with $\gamma = 0$

<table>
<thead>
<tr>
<th>Individual $\phi$/ Temporal</th>
<th>type 1</th>
<th>type 2</th>
<th>type 3</th>
<th>type 4</th>
<th>No Aggreg</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\phi = 0.8$ Flow temporal aggregation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>type 1</td>
<td>12.13</td>
<td>6.96</td>
<td>6.23</td>
<td>4.91</td>
<td>4.19</td>
</tr>
<tr>
<td>type 2</td>
<td>8.54</td>
<td>5.06</td>
<td>4.55</td>
<td>3.61</td>
<td>3.10</td>
</tr>
<tr>
<td>type 3</td>
<td>5.32</td>
<td>3.26</td>
<td>2.94</td>
<td>2.36</td>
<td>2.03</td>
</tr>
<tr>
<td>No Aggreg</td>
<td>2.50</td>
<td>9.95</td>
<td>1.43</td>
<td>1.15</td>
<td>1.00</td>
</tr>
<tr>
<td>$\phi = -0.8$ Stock temporal aggregation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>type 1</td>
<td>65.58</td>
<td>23.72</td>
<td>18.19</td>
<td>8.54</td>
<td>4.19</td>
</tr>
<tr>
<td>type 2</td>
<td>45.98</td>
<td>17.35</td>
<td>13.37</td>
<td>6.31</td>
<td>3.10</td>
</tr>
<tr>
<td>type 3</td>
<td>28.71</td>
<td>11.23</td>
<td>8.69</td>
<td>4.13</td>
<td>2.03</td>
</tr>
<tr>
<td>No Aggreg</td>
<td>13.45</td>
<td>9.95</td>
<td>4.25</td>
<td>2.03</td>
<td>1.00</td>
</tr>
<tr>
<td>$\phi = 0$ Flow temporal aggregation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>type 1</td>
<td>398.45</td>
<td>97.70</td>
<td>95.60</td>
<td>40.23</td>
<td>4.22</td>
</tr>
<tr>
<td>type 2</td>
<td>277.99</td>
<td>70.97</td>
<td>70.04</td>
<td>29.87</td>
<td>3.11</td>
</tr>
<tr>
<td>type 3</td>
<td>173.80</td>
<td>45.75</td>
<td>45.81</td>
<td>19.75</td>
<td>2.04</td>
</tr>
<tr>
<td>No Aggreg</td>
<td>81.67</td>
<td>8.01</td>
<td>22.49</td>
<td>9.80</td>
<td>1.00</td>
</tr>
<tr>
<td>$\phi = -0.8$ Stock temporal aggregation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>type 1</td>
<td>81.96</td>
<td>24.95</td>
<td>22.73</td>
<td>10.67</td>
<td>4.22</td>
</tr>
<tr>
<td>type 2</td>
<td>57.06</td>
<td>18.16</td>
<td>16.62</td>
<td>7.36</td>
<td>3.11</td>
</tr>
<tr>
<td>type 3</td>
<td>35.74</td>
<td>11.74</td>
<td>10.82</td>
<td>5.15</td>
<td>2.04</td>
</tr>
<tr>
<td>No Aggreg</td>
<td>16.73</td>
<td>8.01</td>
<td>5.28</td>
<td>2.52</td>
<td>1.00</td>
</tr>
<tr>
<td>$\phi = 0$ Stock temporal aggregation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>type 1</td>
<td>66.43</td>
<td>24.53</td>
<td>18.59</td>
<td>8.46</td>
<td>4.05</td>
</tr>
<tr>
<td>type 2</td>
<td>47.02</td>
<td>18.07</td>
<td>13.77</td>
<td>6.32</td>
<td>3.03</td>
</tr>
<tr>
<td>type 3</td>
<td>29.24</td>
<td>11.66</td>
<td>8.98</td>
<td>4.36</td>
<td>2.01</td>
</tr>
<tr>
<td>No Aggreg</td>
<td>13.80</td>
<td>21.44</td>
<td>4.41</td>
<td>2.07</td>
<td>1.00</td>
</tr>
</tbody>
</table>

The upper panel shows the results for $\phi = 0.8$, the middle for $\phi = -0.8$ and the bottom for $\phi = 0$. For each sub-panel the upper and middle parts show the average relative efficiency loss when temporal aggregation is flow and stock (individual aggregation is always flow). The bottom part shows the percentage of times that aggregated variances under the temporal stock scheme are smaller than under the temporal flow scheme. No Aggreg column and rows stand for results under temporally disaggregated data (column) and individually disaggregated data (rows).
Note that, as we saw earlier, in the aggregated model the intercept $\mathbf{F}\mu$ becomes individual-specific. The variance of $\mathbf{FU}$ is given by
\[
E(\mathbf{FU}^t\mathbf{F}') = (\mathbf{F}_2\mathbf{F}_2') \otimes (\mathbf{F}_1\Sigma\mathbf{F}_1').
\]
And, since $\Sigma = \sigma_u^2 \mathbf{I}_N + \sigma_\alpha^2 \mathbf{e}_N\mathbf{e}_N'$, it can be re-written as
\[
E(\mathbf{FU}^t\mathbf{F}') = (\mathbf{F}_2\mathbf{F}_2') \otimes \left( \sigma_u^2 \left( \sum_{j=0}^{m-1} w_j \right) \mathbf{I}_\tau + \sigma_\alpha^2 \left( \sum_{j=0}^{m-1} w_j \right)^2 \mathbf{e}_\tau\mathbf{e}_\tau' \right).
\]  (16)

Thus, in the aggregated random effects model the error term presents covariances within individuals different to zero (though the random effects) and heteroskedasticity (through $\mathbf{FU}$).

In order to estimate $\delta$ by least squares we need to weight the observations to obtain a diagonal and homocedastic variance-covariance matrix for the error term. Similarly to estimation with $y_{i,t}$, the appropriate weighting matrix is the square root of the inverse of the variance-covariance matrix of the original aggregated errors, (16), but replacing $\sigma_u^2$, $\sigma_\alpha^2$ and $N$ but their aggregate counterparts given in (16) and $\mathbf{I}_N$ by $(\mathbf{F}_2\mathbf{F}_2')^{-1}$:
\[
\mathbf{V}_\tau^2 = (\mathbf{F}_2\mathbf{F}_2')^{-1} \otimes (\mathbf{F}_1\Sigma\mathbf{F}_1')^{-1} = (\mathbf{F}_2\mathbf{F}_2')^{-1} \otimes \frac{1}{\sigma_u^2(\sum_{j=0}^{m-1} w_j^2)} \left( \mathbf{I}_\tau - \frac{\sigma_\alpha^2(\sum_{j=0}^{m-1} w_j^2)}{\sigma_u^2(\sum_{j=0}^{m-1} w_j^2) + \tau\sigma_\alpha^2(\sum_{j=0}^{m-1} w_j^2)} \mathbf{e}_\tau\mathbf{e}_\tau' \right).
\]

Multiplying both sides of the model by $\mathbf{V}_\tau$:
\[
\mathbf{V}_\tau\mathbf{FY} = \mathbf{V}_\tau\mathbf{FZ}\delta + \mathbf{V}_\tau\mathbf{FU}.
\]

Following the same lines as in (10), this transformation renders $\mathbf{V}_\tau\mathbf{FU}$ homocedastic. The ordinary least squares estimator is
\[
\hat{\delta}^a = (\mathbf{Z}'\mathbf{V}_\tau^2\mathbf{FZ})^{-1}(\mathbf{Z}'\mathbf{V}_\tau^2\mathbf{FY}),
\]
with variance
\[
\text{Var}(\hat{\delta}^a|\mathbf{Z}) = (\mathbf{Z}'\mathbf{V}_\tau^2\mathbf{FZ})^{-1}.
\]  (17)

As in the fixed effects model, variances (15) and (17) are conditional to $\mathbf{Z}$. And, following the same lines, if $\text{Var}(\hat{\delta}|\mathbf{Z}) \leq \text{Var}(\hat{\delta}^a|\mathbf{Z})$ for all $\mathbf{Z}$, then the same dominance applies to the unconditional variances, i.e. $\text{Var}(\hat{\delta}) \leq \text{Var}(\hat{\delta}^a)$.

The following Theorem shows that temporal and individual aggregation increases the unconditional variance of the estimates and thus results in an efficiency loss.

**Theorem 2** In the random effects model (6) with $\gamma = 0$, temporal and individual aggregation results in an efficiency loss of the least squares estimator, i.e. the difference $\text{Var}(\hat{\delta}^a) - \text{Var}(\hat{\delta})$ is semi-positive definite.

As it happened in the fixed effects model, we may wonder if there is a ranking of aggregation schemes in terms of efficiency losses. And as in the fixed effects model, the answer is no. Let $\mathbf{F}_1$ and $\mathbf{F}_1^*$ be two $\tau \times N$ matrices representing two different temporal aggregation schemes. Then, as for the fixed effects model, comparing the efficiency loss of the two aggregation schemes boils down to determine if the matrix
\[
\mathbf{F}_1'(\mathbf{F}_1\Sigma\mathbf{F}_1')^{-1}\mathbf{F}_1 - \mathbf{F}_1^*(\mathbf{F}_1^*\Sigma\mathbf{F}_1^*')^{-1}\mathbf{F}_1^*,
\]  (18)
is semi-positive (negative) definite. Unfortunately this expression is, in general, indefinite. The following Proposition shows that such a ranking is not possible.
Proposition 3 Let the random effects model (6) with \( \gamma = 0 \) and assume two temporal aggregation schemes \( F_1 \) and \( F_1^* \) characterized by the weights \( \omega \) and \( \omega^* \). If there is no constant \( \lambda \in \mathbb{R}\setminus\{0\} \) such that \( \omega \neq \lambda \omega^* \), then (18) is indefinite.

We now turn into individual aggregation. As in the static fixed effects model, comparing two different individual aggregation \( F_2 \) and \( F_2^* \) schemes boils down to determining whether the difference \( P_2 - P_2^* \) is semi positive or negative definite. This is essentially the same problem as in the Proposition 3.

We illustrate the efficiency loss implied by aggregation by means of a simulation study, which has the same structure as the one for the fixed effects model. The differences are that we compare (15) and (17), we set \( \sigma_u^2 = 1 \) and \( \sigma_\alpha^2 = 2 \), and we focus on \( \beta \). Table 2 has the same structure as Table 1 and we extract the same conclusions. In all cases, aggregation entails an important loss of efficiency, even if aggregation is mild, confirming Theorem 2. On the other hand, the comparison of the results for different values of \( \phi \) confirms Proposition 3. No temporal aggregation scheme dominates another. When \( \phi \) is large and positive (negative respectively) flow (stock respectively) temporal aggregation entails less efficiency loss. In fact, from the bottom parts of the sub-panels we conclude that virtually always flow (stock respectively) aggregation does it better when \( \phi \) is large and positive (negative respectively). In the case that the regressor is a white noise, stock or flow aggregation does not make any difference and the proportion of times that one variance is large than other is around 50%.

4.4 Dynamic Models

We now study the impact of aggregation on the parameters \( \theta = (\gamma, \beta) \) in a dynamic linear model. There are many estimation methods available in the literature. Least squares is one of them but it is well known that the least squares estimator for \( \gamma \) has a bias of magnitude \( O(T^{-1}) \) (see e.g. Hsiao (2003)). To avoid this bias, Arellano and Bond (1991) introduced a GMM estimator, henceforth Arellano-Bond. This is the method we choose since it is widely used in the empirical literature. Note that in this case we don’t need to assume the fixed or random effects model (7) and (8) since estimation is performed by taking first differences, and the effects, fixed or random, are swept out.

By contrast to the static model, we study separately the effect of individual and temporal aggregation on the estimators for \( \theta \). Two reasons for this choice. The first is that temporal and individual aggregation have very different consequences on the specification of dynamic linear models, as shown in Section 2. Individual aggregation essentially causes heteroskedasticity while temporal aggregation substantially modifies the structure of the model, inducing moving average terms and dependence with respect to past aggregated exogenous variables. This has important implications in terms of the moment conditions that are used to estimate the parameters. Second, using GMM it is not clear whether estimating with \( \tilde{y}_{a,T} \) instead of with \( y_{i,t} \) implies an efficiency loss. To obtain better insights of this result it is sensible to simplify and look at individual and temporal aggregation separately.

In the sequel we first review the Arellano-Bond estimator. This review is also helpful for introducing further notation. Second, we study the case of individual aggregation. We end with temporal aggregation.

Arellano-Bond is based on the moment conditions

\[
E(y_{i,t-2-j}(u_{i,t} - u_{i,t-1})) = 0, \quad j = 0, 1, 2, \ldots, t - 3
\]

\[
E(x_{i,j}(u_{i,t} - u_{i,t-1})) = 0, \quad j = 1, 2, \ldots, N
\]
## Table 2: Efficiency loss - Model (6) with $\gamma = 0$

<table>
<thead>
<tr>
<th>Individual $\gamma /$ Temporal $\Rightarrow$</th>
<th>type 1</th>
<th>type 2</th>
<th>type 3</th>
<th>type 4</th>
<th>No Aggreg</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\phi = 0.8$ Flow temporal aggregation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>type 1</td>
<td>12.28</td>
<td>6.99</td>
<td>6.26</td>
<td>4.93</td>
<td>4.21</td>
</tr>
<tr>
<td>type 2</td>
<td>8.52</td>
<td>5.08</td>
<td>4.56</td>
<td>3.62</td>
<td>3.11</td>
</tr>
<tr>
<td>type 3</td>
<td>5.29</td>
<td>3.26</td>
<td>2.94</td>
<td>2.36</td>
<td>2.04</td>
</tr>
<tr>
<td>No Aggreg</td>
<td>2.48</td>
<td>9.92</td>
<td>1.43</td>
<td>1.15</td>
<td>1.00</td>
</tr>
<tr>
<td>Stock temporal aggregation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>type 1</td>
<td>52.69</td>
<td>21.93</td>
<td>17.17</td>
<td>8.41</td>
<td>4.21</td>
</tr>
<tr>
<td>type 2</td>
<td>36.81</td>
<td>15.91</td>
<td>12.53</td>
<td>6.18</td>
<td>3.11</td>
</tr>
<tr>
<td>type 3</td>
<td>23.09</td>
<td>10.31</td>
<td>8.15</td>
<td>4.05</td>
<td>2.04</td>
</tr>
<tr>
<td>No Aggreg</td>
<td>10.84</td>
<td>9.92</td>
<td>3.96</td>
<td>1.98</td>
<td>1.00</td>
</tr>
<tr>
<td>% temporal stock $&lt;$ temporal flow</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>type 1</td>
<td>0.15</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>-</td>
</tr>
<tr>
<td>type 2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>-</td>
</tr>
<tr>
<td>type 3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>-</td>
</tr>
<tr>
<td>No Aggreg</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>-</td>
</tr>
<tr>
<td>Stock temporal aggregation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>type 1</td>
<td>52.69</td>
<td>21.93</td>
<td>17.17</td>
<td>8.41</td>
<td>4.21</td>
</tr>
<tr>
<td>type 2</td>
<td>36.81</td>
<td>15.91</td>
<td>12.53</td>
<td>6.18</td>
<td>3.11</td>
</tr>
<tr>
<td>type 3</td>
<td>23.09</td>
<td>10.31</td>
<td>8.15</td>
<td>4.05</td>
<td>2.04</td>
</tr>
<tr>
<td>No Aggreg</td>
<td>10.84</td>
<td>9.92</td>
<td>3.96</td>
<td>1.98</td>
<td>1.00</td>
</tr>
<tr>
<td>% temporal stock $&lt;$ temporal flow</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>type 1</td>
<td>0.15</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>-</td>
</tr>
<tr>
<td>type 2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>-</td>
</tr>
<tr>
<td>type 3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>-</td>
</tr>
<tr>
<td>No Aggreg</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>-</td>
</tr>
<tr>
<td>$\phi = -0.8$ Flow temporal aggregation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>type 1</td>
<td>394.22</td>
<td>96.71</td>
<td>95.34</td>
<td>40.20</td>
<td>4.22</td>
</tr>
<tr>
<td>type 2</td>
<td>275.02</td>
<td>70.41</td>
<td>69.97</td>
<td>29.87</td>
<td>3.11</td>
</tr>
<tr>
<td>type 3</td>
<td>171.66</td>
<td>45.57</td>
<td>45.69</td>
<td>19.74</td>
<td>2.04</td>
</tr>
<tr>
<td>No Aggreg</td>
<td>81.07</td>
<td>8.01</td>
<td>22.41</td>
<td>9.79</td>
<td>1.00</td>
</tr>
<tr>
<td>% temporal stock $&lt;$ temporal flow</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>type 1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>-</td>
</tr>
<tr>
<td>type 2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>-</td>
</tr>
<tr>
<td>type 3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>-</td>
</tr>
<tr>
<td>No Aggreg</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>-</td>
</tr>
<tr>
<td>Stock temporal aggregation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>type 1</td>
<td>64.95</td>
<td>22.69</td>
<td>21.11</td>
<td>10.38</td>
<td>4.22</td>
</tr>
<tr>
<td>type 2</td>
<td>46.00</td>
<td>16.64</td>
<td>15.54</td>
<td>7.67</td>
<td>3.11</td>
</tr>
<tr>
<td>type 3</td>
<td>28.73</td>
<td>10.80</td>
<td>10.11</td>
<td>5.03</td>
<td>2.04</td>
</tr>
<tr>
<td>No Aggreg</td>
<td>13.53</td>
<td>8.01</td>
<td>4.93</td>
<td>2.47</td>
<td>1.00</td>
</tr>
<tr>
<td>% temporal stock $&lt;$ temporal flow</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>type 1</td>
<td>99.56</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>-</td>
</tr>
<tr>
<td>type 2</td>
<td>99.94</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>-</td>
</tr>
<tr>
<td>type 3</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>-</td>
</tr>
<tr>
<td>No Aggreg</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>-</td>
</tr>
<tr>
<td>$\phi = 0$ Flow temporal aggregation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>type 1</td>
<td>66.84</td>
<td>24.56</td>
<td>18.63</td>
<td>8.48</td>
<td>4.05</td>
</tr>
<tr>
<td>type 2</td>
<td>46.60</td>
<td>17.93</td>
<td>13.71</td>
<td>6.31</td>
<td>3.03</td>
</tr>
<tr>
<td>type 3</td>
<td>29.20</td>
<td>11.64</td>
<td>8.96</td>
<td>4.17</td>
<td>2.01</td>
</tr>
<tr>
<td>type 4</td>
<td>13.73</td>
<td>21.43</td>
<td>4.41</td>
<td>2.07</td>
<td>1.00</td>
</tr>
<tr>
<td>Stock temporal aggregation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>type 1</td>
<td>64.43</td>
<td>24.56</td>
<td>18.60</td>
<td>8.45</td>
<td>4.05</td>
</tr>
<tr>
<td>type 2</td>
<td>45.24</td>
<td>17.94</td>
<td>13.71</td>
<td>6.31</td>
<td>3.03</td>
</tr>
<tr>
<td>type 3</td>
<td>28.14</td>
<td>11.58</td>
<td>8.89</td>
<td>4.16</td>
<td>2.01</td>
</tr>
<tr>
<td>No Aggreg</td>
<td>13.27</td>
<td>21.43</td>
<td>4.38</td>
<td>2.06</td>
<td>1.00</td>
</tr>
<tr>
<td>% temporal stock $&lt;$ temporal flow</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>type 1</td>
<td>52.61</td>
<td>50.96</td>
<td>50.71</td>
<td>50.42</td>
<td>-</td>
</tr>
<tr>
<td>type 2</td>
<td>51.79</td>
<td>50.64</td>
<td>50.74</td>
<td>50.05</td>
<td>-</td>
</tr>
<tr>
<td>type 3</td>
<td>53.39</td>
<td>51.09</td>
<td>51.16</td>
<td>51.1</td>
<td>-</td>
</tr>
<tr>
<td>No Aggreg</td>
<td>54.59</td>
<td>52.37</td>
<td>51.2</td>
<td>50.34</td>
<td>-</td>
</tr>
</tbody>
</table>

The upper panel shows the results for $\phi = 0.8$, the middle for $\phi = -0.8$ and the bottom for $\phi = 0$. For each sub-panel the upper and middle parts show the average relative efficiency loss when temporal aggregation is flow and stock (individual aggregation is always flow). The bottom part shows the percentage of times that aggregated variances under the temporal stock scheme are smaller than under the temporal flow scheme. No Aggreg column and rows stand for results under temporally disaggregated data (column) and individually disaggregated data (rows).
This system of moment conditions can be written compactly as $E(q_{i,t}\Delta u_{i,t}) = 0$ where $\Delta = (1 - L)$ and $q_{i,t} = (y_{i,0}, y_{i,1}, ..., y_{i,t-2}, x_{i,1}, ..., x_{i,N})'$ is the vector of instruments. We can gather the moment conditions for the individual $i$ as $E(\Omega_i \Delta u_i) = 0$ where

$$
\Omega_i = \begin{pmatrix}
q_{i,2} & 0 & \ldots & 0 \\
0 & q_{i,3} & \ldots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \ldots & q_{i,N}
\end{pmatrix}
$$

is the matrix of instruments and $\Delta u_i' = (\Delta u_{i,2}, \Delta u_{i,3}, \ldots, \Delta u_{i,N})$. The Arellano-Bond’s estimator $\hat{\theta}$ as

$$
\hat{\theta} = \arg\min (\Omega U') \Psi^{-1} (\Omega U),
$$

where $\Delta U' = (\Delta u_1, \Delta u_2, \ldots, \Delta u_T)$ and $\Omega = (\Omega_1 \Omega_2 \ldots \Omega_T)$. The optimal choice for $\Psi$ is the variance-covariance matrix of $\Omega U$, $\Psi = E(\Omega U U' \Omega')$, which can be approximated by $\hat{\Psi} = \sigma_u^2 (\Omega (I_I \otimes D_N D_N') I')$, where $D_N$ is a $(N - 1) \times N$ matrix of the form

$$
D_N = \begin{pmatrix}
-1 & 1 & \ldots & 0 & 0 \\
0 & -1 & \ldots & 0 & 0 \\
\vdots & \vdots & \ddots & \vdots & \vdots \\
0 & 0 & \ldots & -1 & 1
\end{pmatrix}.
$$

The variance of $\hat{\theta}$ is given by

$$
Var(\hat{\theta}|S, \Omega) = \sigma_u^2 \left( (S' (I_I \otimes D_N') I') (\Omega (I_I \otimes D_N D_N') I')^{-1} (\Omega (I_I \otimes D_N) S) \right)^{-1},
$$

where $S' = (S_1, S_2, \ldots, S_N)$ and $S_i = (y_{i,-1}, x_i)$.

We now turn to the estimation problem when $y_{i,t}$ is only observed aggregated in groups. We apply individual aggregation to the model (GDAS becomes $F_2 \otimes I_N$):

$$(F_2 \otimes I_N) Y = (F_2 \otimes I_N) \alpha + \gamma (F_2 \otimes I_N) Y_{-1} + (F_2 \otimes I_N) X \beta + (F_2 \otimes I_N) U,$$

where $\alpha$ can be fixed or random. The moments conditions are now

$$
E \left( \sum_{i=1}^{I} M_i^a y_{it-2-j} \left( \sum_{i=1}^{I} M_i^a (u_{it} - u_{it-1}) \right) \right) = 0, \quad j = 0, 1, 2, \ldots, t - 3
$$

$$
E \left( \sum_{i=1}^{I} M_i^a x_{ij} \left( \sum_{i=1}^{I} M_i^a (u_{it} - u_{it-1}) \right) \right) = 0, \quad j = 1, 2, \ldots, N
$$

Since Arellano-Bond is robust to the presence of heteroskedasticity, we could estimate $\theta$ immediately. However because the form of the heteroskedasticity is known, it is more appropriate to directly control for it by pre-multiplying the model by the matrix $(F_2 F_2')^{-\frac{1}{2}} \otimes I_N$ that renders the errors homoscedastic:

$$
(F_2 F_2')^{-\frac{1}{2}} F_2 \otimes I_N Y = (F_2 F_2')^{-\frac{1}{2}} F_2 \otimes I_N \alpha + \gamma ((F_2 F_2')^{-\frac{1}{2}} F_2 \otimes I_N) Y_{-1}
$$

$$
+ ((F_2 F_2')^{-\frac{1}{2}} F_2 \otimes I_N) X \beta + ((F_2 F_2')^{-\frac{1}{2}} F_2 \otimes I_N) U.
$$

\footnote{We should include the term $(F_2 \otimes I_N) \mu$ on the right hand side if $\alpha$ is random. For the sake of presentation we continue with the fixed effects model. However, as said above, estimation wise it does not matter if $\alpha$ is fixed or random.}
The individually aggregated model (20) entails a new matrix of instruments, denoted \( \hat{\Omega} \), and that can be shown to be \( \hat{\Omega} = \Omega((F_2F_2')^{-1}F_2 \otimes I_N) \). The moments conditions of the Arellano-Bond estimator are unchanged by individual aggregation, implying that the matrix of instruments of the aggregated model has the same structure as for the disaggregated one.

The Arellano-Bond estimator for \( \theta \) is

\[
\hat{\theta}^a = \text{argmin}(\hat{\Omega}\Delta \hat{U}')(\hat{\Psi}^{-1}(\hat{\Omega}\Delta \hat{U})),
\]
where \( \Delta \hat{U}' = (\Delta u_1, \Delta u_2, \ldots, \Delta u_4) \). The optimal weighting matrix is given by

\[
\hat{\Psi} = E(\hat{\Omega}\Delta U \Delta U' \Omega')
= E(\Omega(P_2 \otimes I_N) \Delta U \Delta \hat{U}'(P_2 \otimes I_N) \Omega'),
\]
which can be approximated by \( \hat{\Psi} = \sigma_u^2(\Omega(P_2 \otimes D_N D_N' \Omega') \). And the variance of \( \hat{\theta}^a \) is given by

\[
Var(\hat{\theta}^a | S, \Omega) = \sigma_u^2 \left( (S'(P_2 \otimes D_N') \Omega' \right) \left( \Omega(P_2 \otimes D_N D_N' \Omega')^{-1} (\Omega(P_2 \otimes D_N) S) \right)^{-1}. (21)
\]

Theorems 1 and 2 in the static model showed that aggregation entails an efficiency loss. In the dynamic case, for temporal aggregation, it is not possible to establish an equivalent Theorem. We know that the estimator based on the disaggregated data is more efficient if

\[
Var(\hat{\theta}|S, \Omega)^{-1} - Var(\hat{\theta}^a|S, \Omega)^{-1}
\]

is semi-positive definite. However this is not always true. To see why, use (19) and (21) to express (22) as

\[
S'C'(CC')^{-1}CS - S'\hat{P}C'(C\hat{P}C')^{-1}C\hat{P}S, (23)
\]
where \( C = \Omega(I_1 \otimes D_N) \) and \( \hat{P} = (P_2 \otimes I_N) \). Expression (23) is the difference between projected variances of \( S \). The first matrix represents the variance that is projected on the column space of the matrix \( C' \), while the second one is the variance that is projected on the column space of the matrix \( \hat{P}C' \). Since the column spaces generated by those two matrices are different, there is no reason for this difference to be semi positive or negative definite.\(^7\)

The fact that (23) can be either semi positive or negative definite, means that we cannot show that

\[
Var(\hat{\theta}^a | S, \Omega) - Var(\hat{\theta}|S, \Omega)
\]
is semi-positive definite for all possible \( S \) and \( \Omega \), which implies that we cannot show that aggregation is unconditionally dominated, i.e. that

\[
Var(\hat{\theta}^a) - Var(\hat{\theta}) (24)
\]
is semi-positive definite.

\(^7\)Note that the preparatory Lemma 1 cannot be used in this context to determine whether (23) is semi positive or semi negative definite. Define \( \varphi_1(C') = C' \), \( \varphi_2(C') = \hat{P}C' \) and \( \psi(C') = C'z \) where \( z \in \mathbb{R}^{N-11N+2} \). The structure of the matrix \( C' \) is such that its columns are not linearly independent, meaning that they cannot be a basis of the column space they generate. Since they are not a basis, a vector lying in the column space of \( C' \) can have many representation in terms of \( z \). This means that the function \( \psi \) is not injective and thus the preparatory Lemma 1 cannot be applied.
As a check, we perform a simple simulation exercise. We simulate 10000 times from (8) for $\beta = 0$ and where $u_{i,t} \sim iid N(0, 1)$, $\alpha_i \sim iid N(0, 1)$, and $\gamma$ takes the values 0.8, 0.4, −0.4 and −0.8. We consider 12 individuals and four different time periods: 4, 8, 10 and 20. For $N = 20$, the GMM estimator should be closed to the least squares estimators and the disaggregated estimator should always deliver smaller variances. Individuals are aggregated as flows (i.e. $M_i^a = 1 \forall a$ and all $i \in a$) and the individual aggregation schemes are the same as in the static cases: i) type 1: $A = 3$, $I_a = 4$, $\forall a$, ii) type 2: $A = 4$, $I_a = 3$, $\forall a$, iii) type 3: $A = 6$, $I_a = 2$, $\forall a$. That is, we aggregate in groups of equal size and the group size decreases with the type. Table 3 shows the simulation results, which are divided in four panels, each for one value of $\gamma$. Each panel is divide in two subpanels. The upper subpanel shows the relative average efficiency loss and the bottom subpanel shows the percentage of times that aggregated variances are smaller than the disaggregated variances. The upper subpanels show that, in average, aggregation entails an information loss. But the subpanels show that it is in fact not possible to establish for GMM an equivalent Theorem for those of least squares, since there are cases in which $\hat{\gamma}^a$ is more efficient than $\gamma$. The proportion of times for which disaggregation is better than aggregation decreases with $\gamma$. As it tends to -1, the proportions tends to 1. And as $N$ increases the disaggregated GMM estimator is always better the aggregated one.

Last, we study the estimation problem when $y_{i,t}$ is observed aggregated in time. We first express the model in the form of (7) or (8) and the apply GDAS:

$$(1 - \gamma^m L^m)FY = \left(\frac{1 - \gamma^m}{1 - \gamma}\right)F\alpha + FX\beta + \cdots + FX_{-m+1}\beta\gamma^{m-1} + \eta,$$

where $\eta = U + \cdots + U_{-m+1}\gamma^{m-1}$. As mention in Section 2, this model has a constrained dynamic response with respect to the exogenous variables and it can be written as:

$$FY = \left(\frac{1 - c_0}{1 - c_0^m}\right)F\alpha + FY_{-m}c_0 + FXc_1 + FX_{-1}c_2 + \cdots + FX_{-m+1}c_m + \eta,$$

where $\tilde{X} = (Y_{-m}, X, \ldots, X_{-m+1})$, $c_0 = \gamma^m$, $c_1 = \beta$, $c_2 = \gamma\beta$, $\ldots$, $c_m = \gamma^{m-1}\beta$, and $c = (c_0, c_1, \ldots, c_m)$. A typical element of (25) is

$$\tilde{y}_{i,T} = \left(\sum_{j=0}^{m-1} c_0^j \right) \left(\sum_{j=0}^{m-1} w_j \right) \alpha_i + c_0 \tilde{y}_{i,T-m} + \tilde{x}_{i,t}c_1 + \cdots + \tilde{x}_{i,t-m+1}c_m + \eta_{i,T},$$

where $c_{m+1} = E(\eta_{a,T}\eta_{a,T-m}) = f(c_0) \neq 0$ (expect for stock aggregation) and $E(\eta_{a,T}\eta_{a,T-sm}) = 0$ for $s = \pm 2, \pm 3, \pm 4, \ldots$. The constraints on the coefficients of the aggregated model can be written in an implicit form as

$$g(c) = \begin{pmatrix}
    c_2 - c_0 \frac{1}{2} \frac{1}{c_1} \\
    c_3 - c_0 \frac{2}{3} \frac{1}{c_1} \\
    \vdots \\
    c_m - c_0 \frac{m-1}{m} \frac{1}{c_1} \\
    c_{m+1} - f(c_0)
\end{pmatrix}.$$
Table 3: Efficiency loss - GMM and Individual Aggregation

<table>
<thead>
<tr>
<th>$\gamma$</th>
<th>N = 4</th>
<th>N = 8</th>
<th>N = 10</th>
<th>N = 20</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>type 1</td>
<td>type 2</td>
<td>type 3</td>
<td>type 3</td>
</tr>
<tr>
<td>0.8</td>
<td>1.24</td>
<td>1.36</td>
<td>1.43</td>
<td>2.26</td>
</tr>
<tr>
<td></td>
<td>1.36</td>
<td>1.76</td>
<td>2.61</td>
<td>3.76</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>% disaggregated &lt; aggregated</td>
<td></td>
</tr>
<tr>
<td></td>
<td>type 1</td>
<td>64.4</td>
<td>82.3</td>
<td>91.16</td>
</tr>
<tr>
<td></td>
<td>type 2</td>
<td>67.4</td>
<td>91.2</td>
<td>97.2</td>
</tr>
<tr>
<td></td>
<td>type 3</td>
<td>67.6</td>
<td>95.1</td>
<td>98.9</td>
</tr>
<tr>
<td>0.4</td>
<td>1.90</td>
<td>1.73</td>
<td>1.79</td>
<td>1.97</td>
</tr>
<tr>
<td></td>
<td>2.61</td>
<td>2.51</td>
<td>2.64</td>
<td>2.96</td>
</tr>
<tr>
<td></td>
<td>3.16</td>
<td>3.35</td>
<td>3.55</td>
<td>3.99</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>% disaggregated &lt; aggregated</td>
<td></td>
</tr>
<tr>
<td></td>
<td>type 1</td>
<td>93.4</td>
<td>99.75</td>
<td>99.9</td>
</tr>
<tr>
<td></td>
<td>type 2</td>
<td>96.1</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>type 3</td>
<td>97.2</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>-0.4</td>
<td>2.08</td>
<td>1.94</td>
<td>1.97</td>
<td>2.01</td>
</tr>
<tr>
<td></td>
<td>3.18</td>
<td>2.97</td>
<td>3.00</td>
<td>3.04</td>
</tr>
<tr>
<td></td>
<td>4.31</td>
<td>4.08</td>
<td>4.08</td>
<td>4.10</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>% disaggregated &lt; aggregated</td>
<td></td>
</tr>
<tr>
<td></td>
<td>type 1</td>
<td>99.8</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>type 2</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>type 3</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>-0.8</td>
<td>2.21</td>
<td>2.09</td>
<td>2.09</td>
<td>2.07</td>
</tr>
<tr>
<td></td>
<td>3.55</td>
<td>3.30</td>
<td>3.28</td>
<td>3.20</td>
</tr>
<tr>
<td></td>
<td>5.07</td>
<td>4.71</td>
<td>4.61</td>
<td>4.41</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>% disaggregated &lt; aggregated</td>
<td></td>
</tr>
<tr>
<td></td>
<td>type 1</td>
<td>99.9</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>type 2</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>type 3</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
</tbody>
</table>

The upper panel shows the results for $\gamma = 0.8$, the second from the top for $\gamma = 0.4$, the second from the bottom for $\gamma = -0.4$ and the bottom for $\gamma = -0.8$. For each sub-panel the upper part shows the average relative efficiency loss when individual aggregation is flow. The bottom part shows the percentage of times that aggregated variances are smaller than the disaggregated variances. All results are under the individual flow scheme.
There are no constraints on \( c_0 \) and \( c_1 \) since they are the parameters to estimate: \( c_0 = \gamma \) and \( c_1 = \beta \). Since the aggregated residuals are autocorrelated of order one at the aggregated frequency, the moment conditions are different from these in the disaggregated case. More specifically, the vector of instruments at frequency \( T \) is \( \mathbf{q}_{i,T}^a = (\bar{y}_{i,0}, \bar{y}_{i,m}, \ldots, \bar{y}_{i,T-2m}, \bar{x}_{i,1}, \ldots, \bar{x}_{i,N})' \). These instruments can be gathered in a matrix

\[
\mathbf{\Omega}_i^a = \begin{pmatrix}
\mathbf{q}_{i,2m}^a & 0 & \ldots & 0 \\
0 & \mathbf{q}_{i,3m}^a & \ldots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \ldots & \mathbf{q}_{i,T}^a
\end{pmatrix}.
\]

Let \( \tilde{\mathbf{\Delta}} \eta_i' = (\tilde{\mathbf{\Delta}} \eta_{i,2m}, \tilde{\mathbf{\Delta}} \eta_{i,3m}, \ldots, \tilde{\mathbf{\Delta}} \eta_{i,T} ) \) where \( \tilde{\mathbf{\Delta}} = (1 - L^m) \). Then the moments conditions for individual \( i \) are

\[
E(\mathbf{\Omega}_i^a \tilde{\mathbf{\Delta}} \mathbf{\eta}_i) = \begin{pmatrix}
E(\mathbf{q}_{i,2m}^a \tilde{\mathbf{\Delta}} \eta_{i,2m}) \\
E(\mathbf{q}_{i,3m}^a \tilde{\mathbf{\Delta}} \eta_{i,3m}) \\
\vdots \\
E(\mathbf{q}_{i,T}^a \tilde{\mathbf{\Delta}} \eta_{i,T})
\end{pmatrix} = \begin{pmatrix}
\mathbf{K}_2(c_{m+1}) \\
\mathbf{K}_3(c_{m+1}) \\
\vdots \\
\mathbf{K}_T(c_{m+1})
\end{pmatrix} = \mathbf{K}(c_{m+1}),
\]

where a typical \( s \) element of \( \mathbf{K}(c_{m+1}) \) is

\[
\mathbf{K}_s(c_{m+1}) = \left( \begin{array}{c}
0, \ldots, 0, c_{m+1}, 0, \ldots, 0
\end{array} \right)'
\]

with the non-zero element corresponding to the autocorrelation of order one in \( \eta \) created by temporal aggregation: \( c_{m+1} = f(c_0) \). For instance, \( E(\mathbf{q}_{i,2m}^a \tilde{\mathbf{\Delta}} \eta_{i,2m}) \) is a system of \( 1 + N \) equations, the first element being \( E(\bar{y}_{i,0} \tilde{\mathbf{\Delta}} \eta_{i,2m}) = E(\bar{y}_{i,0}(\eta_{i,2m} - \eta_{i,m})) \). And this is not zero because \( \eta_{i,m} \) follows and AR(1) process and hence it is related with \( \bar{y}_{i,0} \).

We therefore face a GMM problem with moment conditions that are not zero and with parameter constraints. Let \( \mathbf{\Omega}^a = (\mathbf{\Omega}_1^a, \mathbf{\Omega}_2^a, \ldots, \mathbf{\Omega}_i^a, \ldots) \), \( \tilde{\mathbf{\Delta}} \mathbf{\Upsilon}' = (\tilde{\mathbf{\Delta}} \mathbf{\eta}_1, \tilde{\mathbf{\Delta}} \mathbf{\eta}_2, \ldots, \tilde{\mathbf{\Delta}} \mathbf{\eta}_i) \), and \( \mathbf{\lambda} \) be a vector of Lagrange multipliers. Estimates of \( \mathbf{c} \) and \( \mathbf{\lambda} \) are (see e.g. Gourieroux and Monfort (1995))

\[
(\hat{\mathbf{c}}, \hat{\mathbf{\lambda}}) = \text{argmin}_{\mathbf{c}, \mathbf{\lambda}} \mathcal{L}(\mathbf{c}) + \mathbf{\lambda}'g(\mathbf{c}) \text{ where }
\]

\[
\mathcal{L}(\mathbf{c}) = (\mathbf{\Omega}^a \tilde{\mathbf{\Delta}} \mathbf{\Upsilon} - \mathbf{K}(c_{m+1}))' \hat{\mathbf{\Psi}}^{-1} (\mathbf{\Omega}^a \tilde{\mathbf{\Delta}} \mathbf{\Upsilon} - \mathbf{K}(c_{m+1})),
\]

with optimal weighting matrix is

\[
\hat{\mathbf{\Psi}}^{-1} = E \left( (\mathbf{\Omega}^a \tilde{\mathbf{\Delta}} \mathbf{\Upsilon} - \mathbf{K}(c_{m+1})) (\mathbf{\Omega}^a \tilde{\mathbf{\Delta}} \mathbf{\Upsilon} - \mathbf{K}(c_{m+1}))' \right).
\]

The minimization of (26) is a nonlinear problem with solutions that cannot be found analytically. We know however (see e.g. Gourieroux and Monfort (1995)) the asymptotic variance of \( \sqrt{T^A}(\hat{\mathbf{c}} - \mathbf{c}) \):

\[
\text{Var} \left( \sqrt{T^A}(\hat{\mathbf{c}} - \mathbf{c})|\mathbf{\Omega}^a \right) = (\mathbf{I}_N - \mathbf{M}) J_0^{-1} \mathbf{I}_0 J_0^{-1} (\mathbf{I}_N - \mathbf{M})',
\]

where

\[
\mathbf{M} = J_0^{-1} \frac{\partial g(\mathbf{c})'}{\partial \mathbf{c}} \left( \frac{\partial g(\mathbf{c})}{\partial \mathbf{c}} J_0^{-1} \frac{\partial g(\mathbf{c})'}{\partial \mathbf{c}} \right)^{-1} \frac{\partial g(\mathbf{c})}{\partial \mathbf{c}}
\]

with

\[
J_0 = \lim_{T \to \infty} - \frac{1}{T^A} \frac{\partial^2 \mathcal{L}(\mathbf{c})}{\partial \mathbf{c} \partial \mathbf{c}'} \quad \text{and} \quad \mathbf{I}_0 = \lim_{T \to \infty} \frac{1}{T^A} \frac{\partial \mathcal{L}(\mathbf{c})}{\partial \mathbf{c}} \frac{\partial \mathcal{L}(\mathbf{c})}{\partial \mathbf{c}'}.
\]
Assume that $\gamma > 0$.\(^8\) Then from the estimated parameter vector $\hat{c}$ we can recover the parameters of the disaggregated model: $\hat{\gamma} = \hat{c}_0^1$ and $\hat{\beta} = \hat{c}_1$. Their variances can be backed using the Delta method:

$$Var(\hat{\theta}^a | \Omega^a) = \Pi(c)z'Var(c)z\Pi(c)', \quad (27)$$

where

$$z = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \end{pmatrix}$$

and

$$\Pi(c) = \begin{pmatrix} \frac{1}{m} & \frac{1}{m} & 0 & 0 \\ \frac{1}{m} & \frac{1}{m} & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}.$$

To illustrate the information loss due to temporal aggregation, we consider a simulation exercise with the same set up as the previous one. Time is aggregated as stock and for each $N$ we aggregate every 2, 4, 5 and 10 periods whenever it is possible (e.g. for $N = 8$ we aggregate every 2 and 4 periods but for $N = 10$ we aggregate every 2 and 5 periods). Table 4 shows the simulation results, which are divided in four panels, each for one value of $\gamma$. Each panel is divide in two subpanels. The upper subpanel shows the average relative efficiency loss and the bottom subpanel shows the percentage of times that aggregated variances are smaller than the disaggregated variances. In all cases aggregation entails a loss of efficiency. And the loss can be very large. Two things can explain the presence of the large variances. First, as data are aggregated through time, the correlation between two successive observations decreases. Thus the quality of the instruments used in the GMM deteriorates which leads to an increase in the variance. Second, to compute the variance of the GMM estimators we need to perform two matrix inversions, increasing the probability of numerical instability and the second cause of such large variances.\(^9\) But as $N$ increases the efficiency loss decreases. This is due to the fact that $N$ increases, $\tau$ also increases, improving the accuracy of the estimates. This intuition is also reflected by the fact that as $m$ increases the relative efficiency loss increases.

5 Conclusion

We study the impact of individual and temporal aggregation in linear static and dynamic models for panel data in terms of model specification and efficiency of the estimated parameters. Model wise we find that i) individual aggregation does not affect the model structure, while temporal aggregation may introduce residual autocorrelation, and ii) individual aggregation entails heteroskedasticity while temporal aggregation does not. Estimation wise we find that i) in the static model, estimation by least squares with the aggregated data entails a decrease in the efficiency of the estimated parameters but we cannot rank different aggregation schemes in terms of efficiency, ii) in the dynamic model, estimation by GMM does not necessarily entail a decrease in the efficiency of the estimated parameters under individual aggregation and no analytic comparison can be established for temporal aggregation, though simulation results suggest that temporal aggregation deteriorates the accuracy of the estimates.

Some extensions are possible. Adding more structure to the model, with time-invariant and individual-invariant variables is one. These variables are often found in empirical work to

---

\(^8\)This is an assumption in order to simply the presentation, the next equation could also be rewritten in the case $\gamma < 0$ since the disaggregated model is identified.

\(^9\)We also did the simulation exercise with medians instead of means. In some cases the relative efficiency loss gave more reasonable numbers but not always. Qualitative results however remain unchanged (i.e. aggregation entails an efficiency loss). Results are available under request.
Table 4: Efficiency loss - GMM and Temporal Aggregation

<table>
<thead>
<tr>
<th>$\gamma$</th>
<th>$N = 4$</th>
<th>$N = 8$</th>
<th>$N = 10$</th>
<th>$N = 20$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.8</td>
<td>4.05</td>
<td>2.42</td>
<td>2.09</td>
<td>1.67</td>
</tr>
<tr>
<td>$m = 2$</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>$m = 4$</td>
<td>15 x $10^3$</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>$m = 5$</td>
<td>-</td>
<td>-</td>
<td>26 x $10^4$</td>
<td>6.90</td>
</tr>
<tr>
<td>$m = 10$</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>11 x $10^3$</td>
</tr>
</tbody>
</table>

% disaggregated < aggregated

<table>
<thead>
<tr>
<th>$\gamma$</th>
<th>$N = 4$</th>
<th>$N = 8$</th>
<th>$N = 10$</th>
<th>$N = 20$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.4</td>
<td>6.86</td>
<td>5.10</td>
<td>4.76</td>
<td>4.18</td>
</tr>
<tr>
<td>$m = 2$</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>$m = 4$</td>
<td>32 x $10^3$</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>$m = 5$</td>
<td>-</td>
<td>-</td>
<td>76 x $10^4$</td>
<td>684.6</td>
</tr>
<tr>
<td>$m = 10$</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>57 x $10^6$</td>
</tr>
</tbody>
</table>

% disaggregated < aggregated

<table>
<thead>
<tr>
<th>$\gamma$</th>
<th>$N = 4$</th>
<th>$N = 8$</th>
<th>$N = 10$</th>
<th>$N = 20$</th>
</tr>
</thead>
<tbody>
<tr>
<td>-0.4</td>
<td>16.64</td>
<td>8.09</td>
<td>6.77</td>
<td>4.74</td>
</tr>
<tr>
<td>$m = 2$</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>$m = 4$</td>
<td>49 x $10^3$</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>$m = 5$</td>
<td>-</td>
<td>-</td>
<td>49 x $10^3$</td>
<td>758</td>
</tr>
<tr>
<td>$m = 10$</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>42 x $10^7$</td>
</tr>
</tbody>
</table>

% disaggregated < aggregated

<table>
<thead>
<tr>
<th>$\gamma$</th>
<th>$N = 4$</th>
<th>$N = 8$</th>
<th>$N = 10$</th>
<th>$N = 20$</th>
</tr>
</thead>
<tbody>
<tr>
<td>-0.8</td>
<td>49.10</td>
<td>11.80</td>
<td>6.81</td>
<td>2.92</td>
</tr>
<tr>
<td>$m = 2$</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>$m = 4$</td>
<td>35 x $10^6$</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>$m = 5$</td>
<td>-</td>
<td>-</td>
<td>15.68</td>
<td>5.47</td>
</tr>
<tr>
<td>$m = 10$</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>13 x $10^5$</td>
</tr>
</tbody>
</table>

% disaggregated < aggregated

The upper panel shows the results for $\gamma = 0.8$, the second from the top for $\gamma = 0.4$, the second from the bottom for $\gamma = -0.4$ and the bottom for $\gamma = -0.8$. For each sub-panel the upper part shows the average relative efficiency loss when individual aggregation is flow. The bottom part shows the percentage of times that disaggregated variances are smaller than the aggregated variances. All results are under the temporal stock scheme.
account for time or individual specificities and that may not be captured by the random or fixed effects. Allowing the errors of the disaggregated model to be heteroskedastic is another possibility. However, in models for panel data, the concept of heteroskedasticity is different to the traditional concept in time series, as there may be individual, time or both varying volatility. A last possible extension is to analyze the consequences of aggregation in forecasting.

Appendix

In this final Section we proof the Theorems and Propositions. We start with two preparatory Lemmas.

Preparatory Lemma 1 Let $\mathcal{C}$ be a vectorial space. Consider two projection operators $\pi_1$ and $\pi_2$ defined on $\mathcal{C}$. Let $\mathcal{C}_1^\perp$ and $\mathcal{C}_2^\perp$ be the null spaces associated to the operator $\pi_1$ and $\pi_2$. If $\mathcal{C}_1^\perp \subset \mathcal{C}_2^\perp$ Then $\pi_1 - \pi_2$ is semi-positive definite.

Proof For each $\pi_i$ divide the vectorial space $\mathcal{C}$ as $\mathcal{C} \oplus \mathcal{C}_i^\perp$, where $\oplus$ denotes the direct sum. Consider now the quadratic form $v'(\pi_1 - \pi_2)v$. First rewrite $v$ as $v_2 + v_2^\perp$, where $v_2 \in \mathcal{C}_2$ and $v_2^\perp \in \mathcal{C}_2^\perp$. We have

$$v'(\pi_1 - \pi_2)v = (v_2 + v_2^\perp)'(\pi_1 - \pi_2)(v_2 + v_2^\perp) = (v_2 + v_2^\perp)'\pi_1(v_2 + v_2^\perp) - v_2'^\perp v_2,$$

since $v_2'^\perp\pi_2 = 0$ and $v_2'^\perp\pi_2 = v_2'$. The assumption $\mathcal{C}_1^\perp \subset \mathcal{C}_2^\perp$ implies $\mathcal{C}_2 \subset \mathcal{C}_1$, and thus $\pi_1 v_2 = v_2$ and the above equation can be rewritten as

$$v_2'^\perp\pi_1 v_2 + v_2'^\perp\pi_1 v_2^\perp + v_2'^\perp v_2. \quad (28)$$

Now $v_2^\perp \in \mathcal{C}$, thus we can write $v_2^\perp = z_1 + z_1^\perp$ where $z_1 \in \mathcal{C}_1$ and $z_1^\perp \in \mathcal{C}_1^\perp$ and (28) becomes

$$v_2'^\perp z_1 + z_1'^\perp v_2 + z_1'^\perp z_1. \quad (29)$$

Finally notice that since $z_1 \in \mathcal{C}_2 \cap \mathcal{C}_1$ and $v_2 \in \mathcal{C}_2$, the two first terms of (29) are equal to zero. Thus we can rewrite (29) as $z_1'^\perp z_1 = v_2'^\perp\pi_1 v_2^\perp$. And since $\pi_1$ is a projection matrix, it is semi-positive definite and the Lemma follows.

Proof of Theorem 1 We must show that the difference

$$\sigma^2_u \left( \sum_{j=0}^{m-1} w_j^2 \right) (X'F'V\tilde{Q}_{\tau,A}VFX)^{-1} - \sigma^2_u (X'\tilde{Q}_{N,I}X)^{-1}$$

is semi-positive definite for all $X$. This is equivalent to show that the difference

$$(X'\tilde{Q}_{N,I}X) - \frac{1}{\sum_{j=0}^{m-1} w_j^2} (X'F'V\tilde{Q}_{\tau,A}VFX) \quad (30)$$

is semi-positive definite. Note that

$$F'V\tilde{Q}_{\tau,A}VF = (F_2' \otimes F_1')((F_2F_2')^{-\frac{1}{2}} \otimes I_r)(I_A \otimes Q_\tau)((F_2F_2')^{-\frac{1}{2}} \otimes I_r)(F_2 \otimes F_1)$$

$$= (F_2'(F_2F_2')^{-1}F_2 \otimes F_1'Q_\tau F_1).$$
Let
\[ F_1 = \frac{1}{\sqrt{\sum_{j=0}^{m-1} w_j^2}} F_1 \] and \( P_2 = F_2'(F_2 F_2')^{-1} F_2 \),
where \( F_1 \) is a normalized version of the temporal aggregation matrix and \( P_2 \) is a projection matrix onto the column space of \( F_2' \). Then we can rewrite (30) as
\[
X'((I_1 \otimes Q_N) - (P_2 \otimes \overline{F}_1 Q_{\tau} \overline{F}_1))X = X' \left[ (I_1 - P_2) \otimes Q_N + (P_2 \otimes (Q_N - F_2' Q_{\tau} F_1)) \right] X.
\]

To show that (30) is semi-positive definite, it is sufficient to show that, by the properties of the Kronecker product, all the factors of the Kronecker products in (31) are semi-positive definite.

The matrices \( P_2, I_1 - P_2 \) and \( Q_N \) are semi-positive definite as they are projection matrices. The matrix \( Q_N - \overline{F}_1 Q_{\tau} \overline{F}_1 \) is the difference of two projection matrices (since \( F_1 F_1^T = I_1 \)). By the preparatory Lemma 1, to prove that the difference is semi-positive definite, it is sufficient to show that the null space of \( Q_N, \mathcal{N}^\perp \), is included in the null space of \( \overline{F}_1 Q_{\tau} \overline{F}_1, \mathcal{M}^\perp \), i.e. \( \mathcal{N}^\perp \subseteq \mathcal{M}^\perp \). This can be shown by noting that \( \mathcal{N}^\perp = \{ \lambda e_N \}_{\lambda \in \mathbb{R}} \), the set of all \( N \) dimensional vectors whose entry are all equal to \( \lambda \in \mathbb{R} \). Since
\[
\overline{F}_1 Q_{\tau} \overline{F}_1 \lambda e_N = \lambda \left( \sum_{j=0}^{m-1} w_j^2 \right)^{-1} (I_1 \otimes \omega) Q_{\tau} (I_1 \otimes \omega') e_N
\]
\[
= \lambda \left( \sum_{j=0}^{m-1} w_j^2 \right)^{-1} \left( \left( I_1 \otimes \omega \right)(I_1 \otimes \omega') - \left( I_1 \otimes \omega \right) \frac{1}{\tau} e_\tau e'_\tau (I_1 \otimes \omega') \right) e_N
\]
\[
= \lambda \left( \sum_{j=0}^{m-1} w_j^2 \right)^{-1} (Q_{\tau} \otimes \omega \omega')(e_\tau \otimes e_m)
\]
\[
= \lambda \left( \sum_{j=0}^{m-1} w_j^2 \right)^{-1} (0 \otimes \omega \omega' e_m) = 0,
\]
where the last equality follows since a demeaned vector of ones results in a vector of zeros. Hence \( \mathcal{N}^\perp \subseteq \mathcal{M}^\perp \). \( \square \)

Remark to the proof of Theorem 1 If the temporal aggregation frequency is larger than 2 \((m \geq 2)\) -which always happens if we aggregate temporally-, the inclusion \( \mathcal{N}^\perp \subseteq \mathcal{M}^\perp \) is strict. To see this note that \( \{(e_\tau \otimes z)\}_{z \in \mathbb{R}^m} \in \mathcal{M}^\perp \) but generally \( \notin \mathcal{N}^\perp \)

Proof of Proposition 1 Rewrite (12) as
\[
\left( \sum_{j=0}^{m-1} w_j^2 \right)^{-1} (Q_{\tau} \otimes \omega \omega') - \left( \sum_{j=0}^{m-1} \hat{w}_j^2 \right)^{-1} (Q_{\tau} \otimes \hat{\omega} \hat{\omega}').
\]
Consider now the family of vectors of the form \( z \otimes \omega^c \) where \( z \in \mathbb{R}^\tau \) and \( \omega^c \) is such that \( \omega' \omega^c = 0 \). By construction such vector is in the null space of \( \mathbf{F}'_1 \mathbf{Q}_\tau \mathbf{F}_1 \) since

\[
\mathbf{F}'_1 \mathbf{Q}_\tau \mathbf{F}_1 (z \otimes \omega^c) = \left( \begin{array}{c} \sum_{j=0}^{m-1} w_j^2 \end{array} \right)^{-1} (\mathbf{Q}_\tau \otimes \omega' \omega^c) (z \otimes \omega^c)
\]

\[
= \left( \begin{array}{c} \sum_{j=0}^{m-1} w_j^2 \end{array} \right)^{-1} (\mathbf{Q}_\tau z \otimes \omega' \omega^c) = 0
\]

However generally \( \hat{\omega}' \omega^c \neq 0 \). To see this notice that the condition \( \omega' \omega^c = 0 \) and \( \hat{\omega}' \hat{\omega}^c = 0 \) define two hyperplanes in the \( m \)-dimensional space. Since \( \hat{\omega} \neq \lambda \hat{\omega} \) for all \( \lambda \in \mathbb{R} \setminus \{0\} \) these hyperplanes are not parallel and the result follows. □

**Proof or Proposition 2** If \( \text{rank}(A) > A \) then there is at least one row of \( \mathbf{F}_2 \) that is not a linear combination of the rows in \( \hat{\mathbf{F}}_2 \) and this implies that the columns of \( \mathbf{F}'_2 \) and \( \hat{\mathbf{F}}'_2 \) generate different spaces. □

**Proof of Theorem 2** The proof heavily relies in the proof of Theorem 1. We need to show that

\[
\mathbf{Z}' \mathbf{V}_N^2 \mathbf{Z} - \mathbf{Z}' \mathbf{F}' \mathbf{V}_2^2 \mathbf{F} \mathbf{Z} = \mathbf{Z}' (\mathbf{V}_N^2 - \mathbf{F}' \mathbf{V}_2^2 \mathbf{F}) \mathbf{Z}
\]

is positive definite. We first rewrite

\[
\mathbf{F}' \mathbf{V}_2^2 \mathbf{F} = \mathbf{F}_2'(\mathbf{F}_2 \mathbf{F}_2')^{-1} \mathbf{F}_2 \otimes \mathbf{F}'_1 (\mathbf{F}_1 \Sigma \mathbf{F}_1')^{-1} \mathbf{F}_1.
\]

So

\[
\mathbf{Z}' (\mathbf{V}_N^2 - \mathbf{F}' \mathbf{V}_2^2 \mathbf{F}) \mathbf{Z}
\]

(32)

\[
= \mathbf{Z}' \left[ (\mathbf{I}_1 - \mathbf{P}_2) \otimes \Sigma^{-1} \right] + (\mathbf{P}_2 \otimes (\Sigma^{-1} - \mathbf{F}'_1 (\mathbf{F}_1 \Sigma \mathbf{F}_1')^{-1} \mathbf{F}_1)) \mathbf{Z}.
\]

The proof is done if each component of the Kronecker product is semi-positive definite. The matrix \( \Sigma^{-1} \) is the inverse of a positive definite matrix, and hence it is positive definite matrix. The matrices \( \mathbf{I}_1 - \mathbf{P}_2 \) and \( \mathbf{P}_2 \) are projection matrices, and hence semi-positive definite. Finally

\[
\Sigma^{-1} - \mathbf{F}'_1 (\mathbf{F}_1 \Sigma \mathbf{F}_1')^{-1} \mathbf{F}_1 = \Sigma^{-\frac{1}{2}} \left( \mathbf{I}_1 - \Sigma^{\frac{1}{2}} \mathbf{F}_1 (\mathbf{F}_1 \Sigma \mathbf{F}_1')^{-1} \mathbf{F}_1 \Sigma^{\frac{1}{2}} \right) \Sigma^{-\frac{1}{2}}.
\]

Since the expression within brackets is a projection matrix, the entire matrix product is semi-positive definite. □

**Preparatory Lemma 2** Let \( \mathcal{B}, \mathcal{C} \) and \( \mathcal{D} \) be three vectorial spaces. Consider two linear forms \( \varphi_1 \) and \( \varphi_2 \) such that \( \varphi_i : \mathcal{B} \rightarrow \mathcal{C} \) for \( i = 1, 2 \). Let \( \psi : \mathcal{C} \rightarrow \mathcal{D} \) be an injective linear form. Then

\[
\varphi_1 (\mathcal{B}) \subset \varphi_2 (\mathcal{B}) \Leftrightarrow \psi \circ \varphi_1 (\mathcal{B}) \subset \psi \circ \varphi_2 (\mathcal{B}).
\]
Proof \implies Assume \( \varphi_1(B) \subset \varphi_2(B) \) then by definition have \( \psi \circ \varphi_1(B) \subset \psi \circ \varphi_2(B) \).

\iffalse Since \( \psi \) is injective \( \psi^{-1} \) is unique. Then for each \( v \in \psi \circ \varphi_1(B) \subset \psi \circ \varphi_2(B) \) we have that \( \psi^{-1}(v) \in \varphi_1(B) \). Implying that \( \varphi_1(B) \subset \varphi_2(B) \).
\fi

\( \Longleftarrow \) Assume that \( \psi \circ \varphi_1(B) \subset \psi \circ \varphi_2(B) \). Since \( \psi \) is injective \( \psi^{-1} \) is unique. Then for each \( v \in \psi \circ \varphi_1(B) \subset \psi \circ \varphi_2(B) \) we have that \( \psi^{-1}(v) \in \varphi_1(B) \). Implying that \( \varphi_1(B) \subset \varphi_2(B) \). \( \square \)

Proof of Proposition 3

Start by rewriting (18) as

\[
\Sigma^{-\frac{1}{2}'} \left( \Sigma^{-\frac{1}{2}} \Sigma_{12} \Sigma_{12}' F_1 \Sigma_{12}^{-1} - \Sigma^{-\frac{1}{2}'} \Sigma_{12}' \Sigma_{12}^{-1} \right) \Sigma^{-\frac{1}{2}}.
\]

Since the matrix \( \Sigma^{-\frac{1}{2}'} \) is invertible we only need to consider the matrix difference that is in the bracket. Let \( \varphi_1(z) = F_1' z \) and \( \varphi_2(z) = \hat{F}_1' z \) where \( z \in \mathbb{R}^\tau \) and \( \varphi_1 \) and \( \varphi_2 \) are linear forms mapping \( \mathbb{R}^\tau \) to \( \mathbb{R}^T \). Define \( \psi(Z) = \Sigma^{\frac{1}{2}'} Z \) where \( Z \in \mathbb{R}^{T \times \tau} \). Thus we have that \( \psi \circ \varphi_1 = \Sigma^{\frac{1}{2}} F_1' \) and \( \psi \circ \varphi_2 = \Sigma^{\frac{1}{2}} \hat{F}_1' \). Since \( \Sigma^{\frac{1}{2}} \) is invertible the mapping \( \psi \) is injective and the hypothesis of the preparatory Lemma 2 are fulfilled and hence, from the preparatory Lemma 2, we only need to compare the column space of \( F_1' \) and \( \hat{F}_1' \).

We can assume without loss of generality that the first component of the vector \( \omega \) is non zero. Consider the vector \( F_1' z \) where \( z = (1, 0, \ldots, 0)' \in \mathbb{R}^\tau \). By assumption there is no \( \lambda \in \mathbb{R} \) such that \( \lambda \hat{F}_1' z = F_1' z \). Thus the column space of \( F_1' \) and \( \hat{F}_1' \) do not coincide. \( \square \)

References


