Spatial Aggregation of STAR(1_1) Models

Andrew Gehman* William W. S. Wei†

Abstract
In this paper, we demonstrate that in most cases the spatial aggregate of spatial time series data contains similar properties to the non-aggregate. In particular, we examine the space-time autoregressive model of temporal and spatial lag 1 (the STAR(1_1) model) and through a simulation study see that even when the condition of poolability does not hold, the aggregate can usually be modeled as a STAR(1_1) model with only slightly higher forecast error. The parameters of the aggregate STAR(1_1) model are also derived and shown to be functions of the non-aggregate parameters. Future work will be to generalize these results to other types of STARMA models.

Key Words: Spatial, time series, aggregation, STARMA, poolable

1. Introduction

For decades, time series analysis has been interested in comparing the quality of models and forecasts using aggregate versus non-aggregate data. In forecasting regional unemployment rates, is it beter to use prior regional rates or state rates? How does it impact the analysis if disease counts are known at a state level versus a county level? How much information is lost by using the aggregate data instead of the higher-dimensional non-aggregate data?

In this paper, we deal with spatial aggregation of spatial time series data, although these results could apply to aggregation in other contexts. The space-time autoregressive moving-average (STARMA) modeling approach introducted by Cliff and Ord (1975) and extended by Pfeifer and Deutrch (1980) and Deutsch and Pfeifer (1981) provides a simple understanding of vector time series that are spatially related and which follow the ARMA framework common in time series analysis.

The primary aims of this paper are to discover the model form of the spatial aggregate of a STAR(1_1) model and compare the forecast of the aggregate using either aggregate or non-aggregate data. This will shed light onto problems of modeling and forecasting spatially aggregated data, which is especially relevant when obtaining non-aggregate data is costly or difficult.

1.1 The STAR(1_1) Model

We begin with the mean-stationary, mean-centered, non-aggregate STAR(1_1) time series as defined by Pfeifer and Deutrch (1980) and Deutsch and Pfeifer (1981) for $t = 1, 2, ..., T$. Each of the r variables correspond to locations which are arranged on a 2-dimensional plane. At time t,

$$\begin{aligned}\mathbf{Z}_t &= \phi_{z,0}\mathbf{Z}_{t-1} + \phi_{z,1}\mathbf{W}_z\mathbf{Z}_{t-1} + \mathbf{a}_t \\ &= [\phi_{z,0}\mathbf{I}_r + \phi_{z,1}\mathbf{W}_z]\mathbf{Z}_{t-1} + \mathbf{a}_t \\ &= \mathbf{B}\mathbf{Z}_{t-1} + \mathbf{a}_t, \end{aligned} \tag{1}$$

which is a special case of the VAR(1) model, with $\mathbf{B} = \phi_{z,0}\mathbf{I}_r + \phi_{z,1}\mathbf{W}_z$. The parameters $\phi_{z,0}$ and $\phi_{z,1}$ are scalars such that $|\phi_{z,0}| + |\phi_{z,1}| < 1$ to ensure stationarity (Pfeifer and

*Temple University, 1801 Liacouras Walk, Philadelphia, PA, 19122
†Temple University, 1801 Liacouras Walk, Philadelphia, PA, 19122

Jay Deutsch, 1980). We think of $\phi_{z,0}$ as the purely autoregressive parameter and $\phi_{z,1}$ as the spatial parameter, with each being the same for all locations.

In addition, the error vectors $\mathbf{a}_t$ are independent and distributed as $N(\mathbf{0}, \mathbf{\Sigma})$, where $\mathbf{\Sigma}$ is any non-singular matrix, following the notation of Deutsch and Pfeifer (1981) which allows for differences in the error variances as well as their contemporaneous correlation. Thus, we have $E[\mathbf{a}_t\mathbf{a}_t^{'}] = \mathbf{\Sigma}$, and $E[\mathbf{a}_t\mathbf{a}_{t+k}^{'}] = \mathbf{0}$ for $k > 0$. In addition, we define the time lag-k covariance matrix of of our time series variable as $\mathbf{\Gamma}_z(k) = E[\mathbf{Z}_t\mathbf{Z}_{t+k}^{'}]$ for $k \geq 0$.

In this paper, we limit our scope to modeling spatial dependence between first neighbors, defined as locations sharing a common border. Thus, $\mathbf{W}_z$ consists of elements $[\mathbf{W}_z]_{(i,j)}$, which equal $1/n_i$ when locations i and j share a common border, and 0 otherwise. Here, n_i denotes the number of bordering neighbors that location i has, row-standardizing the matrix such that $\mathbf{W}_z\mathbf{1} = \mathbf{1}$, where $\mathbf{1}$ is the column vector of all ones. Thus each location within $\mathbf{Z}_t$ is modeled in terms of the lagged values of its first neighbors as well as of its own lagged value.

Then, a qxr aggregation matrix $\mathbf{A}$ can be used to construct a new set of q $(< r)$ time series, $\mathbf{Y}_t$ as the spatial aggregate of $\mathbf{Z}_t$, such that $\mathbf{Y}_t = \mathbf{A}\mathbf{Z}_t$. We consider the case where $\mathbf{A}$ is binary, with each element $[\mathbf{A}]_{(i,j)}$ taking the value 1 when $Z_{j,t}$ is included in the aggregate $Y_{i,t}$, and 0 otherwise. Thus, the elements of row i in $\mathbf{A}$ construct $Y_{i,t}$ as the sum of designated elements of $\mathbf{Z}_t$. This aggregation procedure is applicable for data that are meaningful when summed, such as sales data, the number of unemployed people, or the number of reports of a disease or crime. Other forms of $\mathbf{A}$ could be adopted for different types of data and purposes.

1.2 Literature Review on Time Series Aggregation

Many have explored the result of aggregating a vector time series into a univariate aggregate (Wei and Abraham, 1981; Kohn, 1982; Pino et al., 1987; Hendry and Hubrich, 2006, 2011), and into a multivariate aggregate Lutkepohl (1984, 1987, 2009). The work of Lutkepohl (1987) provided a wealth of information on forecasting linear transformations of VARMA time series using a general qxr aggregation matrix. He showed that the aggregate also follows a VARMA model and gives bounds for its AR and MA orders. These bounds are considerably wide, so his results provide little guidance for our model.

Much of the relevent work in Kohn (1982) and Lutkepohl (1987) focused on the concept of poolability. The poolability condition for a r-dimensional non-aggregate VAR(1) time series and its q-dimensional aggregate VAR(1) is $\mathbf{AB} = \mathbf{DA}$, where $\mathbf{B}$ is the rxr parameter matrix of the non-aggregate VAR(1) and $\mathbf{D}$ is the qxq parameter matrix of the aggregate VAR(1). Implied in this condition is that the aggregate of a poolable time series follows the same AR order as the non-aggregate. The main result of the poolability condition is that the forecasts of the aggregate using past aggregate information are equivalent to those obtained using non-aggregate information.

Kohn (1982) and more recently Hendry and Hubrich (2006) and Hendry and Hubrich (2011) developed a test to determine poolability in practice for the case of a univariate aggregate ($q = 1$). The aggregate model was fit by the aggregate information and a set of lagged non-aggregates. The test for poolability was thus a test for the significance of the non-aggregate terms added in this model (testing to see if the aggregate information is sufficient to model the aggregate). Bun (2004) provided a similar test and asymptotic results for time series of seemingly unrelated regression equations. Kohn mentioned but never explored testing for poolability in the case of a multivariate aggregate.

Two papers have tested the best method of forecasting aggregates from spatially-dependent time series following STAR(1_1) models through simulations. The driving question is whether it is best to forecast the aggregate directly or aggregate the forecast using non-aggregate information. Giacomini and Granger (2004) compared four methods for forecasting the univariate aggregate: (1) from a univarate ARMA model of the aggregate, (2) from separate ARMA models of each non-aggregate, (3) from a VAR(1) model of the non-aggregates, and (4) from a STAR(1_1) model of the non-aggregates. They demonstrated that in scenarios when the poolability condition held for the data generating process, it was best to forecast the univariate aggregate directly (1), as it was parsimonious and efficient. In all other cases, the non-aggregate STAR(1_1) method (4) outperformed the other methods, as it capture the spatial-correlation and was also parsimonious.

Lastly, Arbia et al. (2010) extended Giacomini and Granger's study by considering a multivariate aggregate of a STAR(1_1) model. They compared the multivariate forecasts using three methods: (1) from separate ARMA models of each aggregate, (2) from a VAR(1) model of the aggregate, and (3) from a STAR(1_1) model of the aggregate. Again, the aggregate STAR(1_1) model (3) had the best-performing forecasts. No justification is made for using a STAR(1_1) model to fit the multivariate aggregate of a STAR(1_1) model. They do state that the univariate aggregate is approximately AR(1) if the the column sums of the weighting matrix $\mathbf{W}_z$ are similar. This is just a restatement of the poolability condition in the STAR(1_1) case, and it suggests approximate poolability.

In summary, these reviewed papers provide the condition of poolability using the non-aggregate and aggregate parameter matrices and the aggregation matrix, whereby the multivariate aggregate of a poolable VAR(1) model is known to follow a VAR(1) model and has forecasts equivalent to the non-aggregate model. Many results focus on the univariate aggregate, and few have explored the special properties offered by the STAR(1_1) model. More importantly, no poolability test for the multivariate aggregate exists in literature.

In this paper, we extend these studies in several ways. First, we provide poolability results for the STAR(1_1) model, including aggregate parameter derivations. We also examine the validity of fitting the aggregate with a STAR(1_1) model when poolability does not hold. Second, we carry out simulation comparisons of approximate poolability for the spatial aggregate of a STAR(1_1) model under varying instances of the error covariance, the non-aggregate parameters, and the aggregation scheme. We look to see when it reasonable to model the aggregate with a STAR(1_1) model and how that model's one-step ahead forecasts compare to those from the non-aggregate model. Finally, we conclude with a summary of our results and direction for future work.

2. The Spatial Aggregate of a STAR(1_1) Model

2.1 Poolability for the STAR(1_1) Model

The poolability condition for VAR(1) model is $\mathbf{AB} = \mathbf{DA}$, where $\mathbf{B}$ and $\mathbf{D}$ are the parameter matrices in the non-aggregate and aggregate models respectively. For the STAR(1_1) model which is a special case of the VAR(1), the poolability condition is the same, where $\mathbf{B} = \phi_{z,0}\mathbf{I}_r + \phi_{z,1}\mathbf{W}_z$ and $\mathbf{D} = \phi_{y,0}\mathbf{I}_q + \phi_{y,1}\mathbf{W}_y$. We now prove results for poolability in the STAR(1_1) model.

Theorem 1 *Given a STAR(1_1) model as in (1), the following are true*

(i) The poolability condition is equivalent to $\mathbf{AW}_z = \psi_1\mathbf{A} + \psi_2\mathbf{W}_y\mathbf{A}$, where $\psi_1 = (\phi_{y,0} - \phi_{z,0})/\phi_{z,1}$ and $\psi_2 = \phi_{y,1}/\phi_{z,1}$.

(ii) Under poolability, the aggregate variable is a STAR(1_1) model.

(iii) Under poolability, the aggregated error and one-step ahead forecasts of the non-aggregate model are equivalent to those of the aggregate model.

Proof 1 *As a proof of these results,*

(i) The poolability conditions implies that:

$$\begin{aligned}\mathbf{0} &= \mathbf{AB} - \mathbf{DA} \\ &= \mathbf{A}[\phi_{z,0}\mathbf{I}_r + \phi_{z,1}\mathbf{W}_z] - [\phi_{y,0}\mathbf{I}_q + \phi_{y,1}\mathbf{W}_y]\mathbf{A} \\ &= \phi_{z,0}\mathbf{A} + \phi_{z,1}\mathbf{AW}_z - \phi_{y,0}\mathbf{A} - \phi_{y,1}\mathbf{AW}_y \\ &= \phi_{z,1}[\mathbf{AW}_z - \frac{\phi_{y,0} - \phi_{z,0}}{\phi_{z,1}}\mathbf{A} - \frac{\phi_{y,1}}{\phi_{z,1}}\mathbf{AW}_y] \\ &= \phi_{z,1}[\mathbf{AW}_z - \psi_1\mathbf{A} - \psi_2\mathbf{W}_y\mathbf{A}], \\ &\qquad where \quad \psi_1 = (\phi_{y,0} - \phi_{z,0})/\phi_{z,1} \quad and \quad \psi_2 = \phi_{y,1}/\phi_{z,1}.\end{aligned}$$

Thus, $\mathbf{AW}_z = \psi_1\mathbf{A} + \psi_2\mathbf{W}_y\mathbf{A}$.

(ii) Using (1) and Theorem 1 Part (i),

$$\begin{aligned}\mathbf{Y}_t &= \mathbf{AZ}_t \\ &= \phi_{z,0}\mathbf{AZ}_{t-1} + \phi_{z,1}\mathbf{AW}_z\mathbf{Z}_{t-1} + \mathbf{Aa}_t \\ &= \phi_{z,0}\mathbf{Y}_{t-1} + \phi_{z,1}[\psi_1\mathbf{A} + \psi_2\mathbf{W}_y\mathbf{A}]\mathbf{Z}_{t-1} + \mathbf{e}_t \\ &= \phi_{z,0}\mathbf{Y}_{t-1} + \phi_{z,1}(\phi_{y,0} - \phi_{z,0})/\phi_{z,1}\mathbf{AZ}_{t-1} + \phi_{z,1}\phi_{y,1}/\phi_{z,1}\mathbf{W}_y\mathbf{AZ}_{t-1} + \mathbf{e}_t \\ &= \phi_{z,0}\mathbf{Y}_{t-1} + (\phi_{y,0} - \phi_{z,0})\mathbf{Y}_{t-1} + \phi_{y,1}\mathbf{W}_y\mathbf{Y}_{t-1} + \mathbf{e}_t \\ &= \phi_{y,0}\mathbf{Y}_{t-1} + \phi_{y,1}\mathbf{W}_y\mathbf{Y}_{t-1} + \mathbf{e}_t, \qquad (2)\end{aligned}$$

where $\mathbf{e}_t = \mathbf{Aa}_t$ *is independent and distributed as* $N(\mathbf{0}, \mathbf{A\Sigma A}')$. *Also, as a result,* $E[\mathbf{e}_t\mathbf{e}_{t+k}'] = E[\mathbf{Aa}_t\mathbf{a}_{t+k}'\mathbf{A}'] = \mathbf{A}E[\mathbf{a}_t\mathbf{a}_{t+k}']\mathbf{A}' = \mathbf{0}$, *for* $k > 0$. *Thus, (2) is a STAR*(1_1) *model.*

(iii) This follows as the consequence of the error term $\mathbf{a}_t$ *in the non-aggregate model and the error term* $\mathbf{e}_t$ *in the aggregate model is related by* $\mathbf{e}_t = \mathbf{Aa}_t$ *under poolability.* ■

We recognize that for most configurations of $\mathbf{W}_z$, $\mathbf{W}_y$, and $\mathbf{A}$ there will not exist values of ψ_1 and ψ_2 that satisfy the poolability condition in Part (i) of Theorem 1. One exception is when there are $r = 4$ non-aggregate locations arranged on a 2x2 grid aggregated into $q = 2$ aggregate locations. In this case,

$$\mathbf{W}_z = \begin{bmatrix} 0 & 1/2 & 1/2 & 0 \\ 1/2 & 0 & 0 & 1/2 \\ 1/2 & 0 & 0 & 1/2 \\ 0 & 1/2 & 1/2 & 0 \end{bmatrix},$$

$$\mathbf{A} = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{bmatrix},$$

$\mathbf{W}_y = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$, and thus,

$\mathbf{AW}_z = \begin{bmatrix} 1/2 & 1/2 & 1/2 & 1/2 \\ 1/2 & 1/2 & 1/2 & 1/2 \end{bmatrix} = \begin{bmatrix} 1/2 & 1/2 & 0 & 0 \\ 0 & 0 & 1/2 & 1/2 \end{bmatrix} + \begin{bmatrix} 0 & 0 & 1/2 & 1/2 \\ 1/2 & 1/2 & 0 & 0 \end{bmatrix} = 1/2\mathbf{A} + 1/2\mathbf{W}_y\mathbf{A}$, for $\psi_1 = 1/2$ and $\psi_1 = 1/2$. Thus, $\phi_{y,0} = \phi_{z,0} + 1/2\phi_{z,1}$ and $\phi_{y,1} = 1/2\phi_{z,1}$ according to Theorem 1 Part (i).

2.2 Parameters of the Poolable Aggregate STAR(1_1) Model

As mentioned previously, poolability holds when $\mathbf{AW}_z - \psi_1\mathbf{A} - \psi_2\mathbf{W}_y\mathbf{A} = \mathbf{0}$. One way to solve for ψ_1 and ψ_2 (and thus to derive $\phi_{y,0}$ and $\phi_{y,1}$) is to minimize the difference in $\mathbf{C} = \mathbf{AB} - \mathbf{DA} = \phi_{z,1}[\mathbf{AW}_z - \psi_1\mathbf{A} - \psi_2\mathbf{W}_y\mathbf{A}]$, such that it equals $\mathbf{0}$. We can do this through least squares estimation.

Theorem 2 *Given a poolable STAR(1_1) model in (1), the parameters of the aggregate STAR(1_1) model are* $\phi_{y,0} = \phi_{z,0} + \frac{tr(\mathbf{A}'\mathbf{AW}_z)}{tr(\mathbf{A}'\mathbf{A})}\phi_{z,1}$ *and* $\phi_{y,1} = \frac{tr(\mathbf{A}'\mathbf{W}_y'\mathbf{AW}_z)}{tr(\mathbf{A}'\mathbf{W}_y'\mathbf{W}_y\mathbf{A})}\phi_{z,1}$.

Proof 2 *The equation* $\mathbf{AW}_z = \psi_1\mathbf{A} + \psi_2\mathbf{W}_y\mathbf{A} + \mathbf{C}/\phi_{z,1}$ *has* $\mathbf{C}/\phi_{z,1}$ *as its error and is given by*

$$\begin{aligned} vec(\mathbf{AW}_z) &= \psi_1 vec(\mathbf{A}) + \psi_2 vec(\mathbf{W}_y\mathbf{A}) + vec(\mathbf{C}/\phi_{z,1}) \\ &= \begin{bmatrix} vec(\mathbf{A}) & vec(\mathbf{W}_y\mathbf{A}) \end{bmatrix} \begin{bmatrix} \psi_1 \\ \psi_2 \end{bmatrix} + vec(\mathbf{C}/\phi_{z,1}). \end{aligned}$$

Then, the ordinary least squares solution is

$$\begin{aligned} \begin{bmatrix} \psi_1 \\ \psi_2 \end{bmatrix} &= \begin{bmatrix} vec(\mathbf{A})'vec(\mathbf{A}) & vec(\mathbf{A})'vec(\mathbf{W}_y\mathbf{A}) \\ vec(\mathbf{W}_y\mathbf{A})'vec(\mathbf{A}) & vec(\mathbf{W}_y\mathbf{A})'vec(\mathbf{W}_y\mathbf{A}) \end{bmatrix}^{-1} \begin{bmatrix} vec(\mathbf{A})'vec(\mathbf{AW}_z) \\ vec(\mathbf{W}_y\mathbf{A})'vec(\mathbf{AW}_z) \end{bmatrix} \\ &= \begin{bmatrix} tr(\mathbf{A}'\mathbf{A}) & tr(\mathbf{A}'\mathbf{W}_y\mathbf{A}) \\ tr(\mathbf{A}'\mathbf{W}_y'\mathbf{A}) & tr(\mathbf{A}'\mathbf{W}_y'\mathbf{W}_y\mathbf{A}) \end{bmatrix}^{-1} \begin{bmatrix} tr(\mathbf{A}'\mathbf{AW}_z) \\ tr(\mathbf{A}'\mathbf{W}_y'\mathbf{AW}_z) \end{bmatrix} \\ &= \begin{bmatrix} tr(\mathbf{A}'\mathbf{A}) & \mathbf{0} \\ \mathbf{0} & tr(\mathbf{A}'\mathbf{W}_y'\mathbf{W}_y\mathbf{A}) \end{bmatrix}^{-1} \begin{bmatrix} tr(\mathbf{A}'\mathbf{AW}_z) \\ tr(\mathbf{A}'\mathbf{W}_y'\mathbf{AW}_z) \end{bmatrix} \\ &= \begin{bmatrix} \frac{tr(\mathbf{A}'\mathbf{AW}_z)}{tr(\mathbf{A}'\mathbf{A})} \\ \frac{tr(\mathbf{A}'\mathbf{W}_y'\mathbf{AW}_z)}{tr(\mathbf{A}'\mathbf{W}_y'\mathbf{W}_y\mathbf{A})} \end{bmatrix}. \end{aligned}$$

This is due to $tr(\mathbf{A}'\mathbf{W}_y\mathbf{A}) = tr(\mathbf{AA}'\mathbf{W}_y) = \sum_{i,j}(\mathbf{AA}')_{(i,j)} \circ (\mathbf{W}_y)_{(i,j)} = 0$*, since* $(\mathbf{AA}')_{(i,j)} = 0$ *for* $i \neq j$ *and* $(\mathbf{W}_y)_{(i,j)} = 0$ *for* $i = j$*, and similarly for* $tr(\mathbf{A}'\mathbf{W}_y'\mathbf{A})$*. By Part (i) of Theorem 1,* $\phi_{y,0} = \phi_{z,0} + \psi_1\phi_{z,1}$ *and* $\phi_{y,1} = \psi_2\phi_{z,1}$. ■

See that all components of these derivations are determined by the non-aggregate model and the aggregation scheme, and don't depend on $\mathbf{\Sigma}$. In addition, the example of poolability at the end of Section 2.1 can be demonstrated to follow Theorem 2. Here, $\psi_1 = \frac{tr(\mathbf{A}'\mathbf{AW}_z)}{tr(\mathbf{A}'\mathbf{A})} = \frac{2}{4} = \frac{1}{2}$ and $\psi_2 = \frac{tr(\mathbf{A}'\mathbf{W}_y'\mathbf{AW}_z)}{tr(\mathbf{A}'\mathbf{W}_y'\mathbf{W}_y\mathbf{A})} = \frac{2}{4} = \frac{1}{2}$. This agrees with the values of ψ_1 and ψ_2 in the example earlier.

2.3 The Aggregate Model without Poolability

In the case where the poolability condition in Section 2.1 does not hold, the assumption of the STAR(1_1) model for the aggregate may no longer be valid. Without poolability, fitting the aggregate as a STAR(1_1) model results in

$$\begin{aligned}\mathbf{Y}_t &= \mathbf{ABZ}_{t-1} + \mathbf{Aa}_t \\ &= \mathbf{DAZ}_{t-1} + (\mathbf{AB} - \mathbf{DA})\mathbf{Z}_{t-1} + \mathbf{Aa}_t \\ &= \mathbf{DY}_{t-1} + \mathbf{CZ}_{t-1} + \mathbf{Aa}_t \\ &= \mathbf{DY}_{t-1} + \mathbf{v}_t \\ &= [\phi_{y,0}\mathbf{I}_q + \phi_{y,1}\mathbf{W}_y]\mathbf{Y}_{t-1} + \mathbf{v}_t \\ &= \phi_{y,0}\mathbf{Y}_{t-1} + \phi_{y,1}\mathbf{W}_y\mathbf{Y}_{t-1} + \mathbf{v}_t. \end{aligned} \tag{3}$$

The error term is now $\mathbf{v}_t = \mathbf{CZ}_{t-1} + \mathbf{Aa}_t$, where $\mathbf{C} = \mathbf{AB} - \mathbf{DA} = \phi_{z,1}[\mathbf{AW}_z - \psi_1\mathbf{A} - \psi_2\mathbf{W}_y\mathbf{A}]$. The term $\mathbf{CZ}_{t-1}$ represents additional error or misspecification due to the loss of informtion in only using the lagged aggregate. So $E[\mathbf{v}_t\mathbf{v}_t^{'}] = \mathbf{C\Gamma}_z(0)\mathbf{C}^{'} + \mathbf{A\Sigma A}^{'} = \mathbf{H}$, and $E[\mathbf{v}_t\mathbf{v}_{t+k}^{'}] = \mathbf{C\Gamma}_z(k)\mathbf{C}^{'} + \mathbf{A\Sigma}(\mathbf{B}^{'})^{k-1}\mathbf{C}^{'} \neq \mathbf{0}$, for $k > 0$. Here, $\mathbf{\Gamma}_z(k) = E[\mathbf{Z}_t\mathbf{Z}_{t+k}^{'}]$ for $k \geq 0$.

This is an obvious case of serial autocorrelation in the error term. However, as would be the case when $\mathbf{C}$ is small and close to $\mathbf{0}$, the error term then may be close to white noise. We propose that in many cases $\mathbf{C} \approx \mathbf{0}$ by the minimizing methods of Theorem 2, which results in the adequacy of the STAR(1_1) model in (3). This would be considered approximate poolability.

2.4 Testing Approximate Poolability in the STAR(1_1) Model

To test this about $\mathbf{v}_t$, the multivariate Portmanteau test is used. The statistic is

$$P(m) = T^2 \sum_{k=1}^{m} (T-k)^{-1} tr(\hat{\Gamma}_v^{'}(k)\hat{\Gamma}_v^{-1}(0)\hat{\Gamma}_v(k)\hat{\Gamma}_v^{-1}(0)), \tag{4}$$

where $\hat{\Gamma}_v(k) = \frac{1}{T}\sum_{t=k+1}^{T} \hat{\mathbf{v}}_t\hat{\mathbf{v}}_{t-k}^{'}$ (see Ljung and Box (1978); Lutkepohl (2009)). The statistic can be compared with its known distribution, a chi-squared with $r^2(m-1)$ degrees of freedom. If the null hypothesis of error white noise is not rejected, the model in (3) is a STAR(1_1) model and is deemed adequate for modeling the multivariate aggregate.

To evaluate the fit and forecast of the aggregate model, it can be compared with the aggregate of the non-aggregate model. In practice, both (1) and (3) can be fit by feasible generalized least squares (FGLS). By this method, the error covariance matrices $\mathbf{\Sigma}$ and $\mathbf{H}$ are estimated from the residuals of an ordinary least squares (OLS) fit and then used to estimate the model parameters.

To measure the forecast accuracy, a series of one-step ahead forecast errors from both models will be compared using an out-of-sample test set of size T^*. The estimated residuals for the out-of-sample time points will be used to compute the mean squared forecast error (MSFE). The MSFE of the aggregated forecasts of the non-aggregate model is

$$MSFE_{(z)} = \frac{1}{T^*} \sum_{t=T+1}^{T+T^*} \hat{\mathbf{a}}_t^{'}\mathbf{A}^{'}\mathbf{A}\hat{\mathbf{a}}_t, \tag{5}$$

while the MSFE direct from the aggregate model is

$$MSFE_{(y)} = \frac{1}{T^*} \sum_{t=T+1}^{T+T^*} \hat{\mathbf{v}}_t^{'}\hat{\mathbf{v}}_t. \tag{6}$$

3. Multivariate Aggregate Simulation

In order to determine under what conditions (3) is a STAR(1_1) and the aggregate model and the non-aggregate model have comparable forecast error, we designed a simulation experiment in which we varied the non-aggregate error covariance matrix, non-aggregate parameters, and level of aggregation.

3.1 Simulation Details

First, we considered the locational arrangement where the $r = 16$ locations were arranged on a 4x4 lattice grid. The locations are numbered going left to right then top to bottom. Having 16 locations allows for a balance of corner, edge, and middle locations, minimizing the impact of edge effects. For a discussion of edge effects see Anselin (1988) or Griffith (1988).

Second, we consider three cases of the error covariance matrix:

- Independent errors with equal variance, $\sigma^2 = 1$ (IND.)
- Moderate correlation of error with varying variances, such that the errors for first neighbors had a correlation of 0.5, those of second neighbors was 0.3, and all others had a correlation of 0.1; with the errors for locations 1 through 16 in having variances 16 through 1, respectively (MOD.)
- Large positive correlation of errors ($\rho = 0.9$) between all locations with equal variance, $\sigma^2 = 1$ (HIGH)

Third, we had the two STAR(1_1) parameters ranged from 0.1 to 0.8 by increments of 0.45 in cases satisfying $|\phi_{z,0}| + |\phi_{z,1}| < 1$ for purposes of stationarity, resulting in 6 combinations of $(\phi_{z,0}, \phi_{z,1})$: (0.10,0.10), (0.45,0.10), (0.80,0.10), (0.10,0.45), (0.45,0.45), and (0.10,0.80).

Finally, we conducted the same simulation experiment as above with $q = 8, 4, 2,$ and 1. In each, the first r/q locations will be aggregated together to form aggregate location 1, the second r/q locations will form aggregate location 2, and so on. It should be noted that in none of these cases does poolability hold as in Theorem 1. That is, for these $\mathbf{W}_z$, $\mathbf{A}$, and $\mathbf{W}_y$ there is no combination of ψ_1 and ψ_2 that makes the poolability condition true.

The simulation was performed for $N = 250$ times under each of the 36 combinations of $\mathbf{\Sigma}$ and $\mathbf{A}$ mentioned above and for all 6 parameter combinations. Within each, the STAR(1_1) time series of length 301 was simulated according to the model in (1), and then the first observation was discarded, so $T + T^* = 300$. The first $T = 200$ observations were used to fit the model, and the final $T^* = 100$ were preserved as an out-of-sample forecast set. The non-aggregate model (1) and aggregate model (3) were fit with OLS, and their residuals were used to estimate the model parameters via FGLS. Iin addition, the aggregate STAR(1_1) residuals were tested using the Portmanteau test in (4) to see if they were white noise. Finally, using the FGLS parameter estimates, the mean squared forecast error (MSFE) of the one-step ahead forecast errors from both models were calculated as in (5) and (6). The results are given below.

3.2 Simulation Results

In Table 1, we examine the three main questions: 1) How often the residuals of the aggregate STAR(1_1) model white noise? 2) How often does the aggregate STAR(1_1) model provide better forecasts of the aggregate than the non-aggregate STAR(1_1) model? and 3)

What percent increase in MSFE is caused by forecasting the aggregate with the aggregate STAR(1_1) model over the non-aggregate STAR(1_1) model?

From Table 1, we see that in most cases, the STAR(1_1) fit of the aggregate is adequate a majority (80 to 90%) of the time. In addition, the aggregate model produced better forecast results than the higher dimensional non-aggregate model around 30 to 45% of the time for most configurations. Lastly, the mean percent increase in MSFE caused by using the lower dimension aggregate model over the non-aggregate model was under or around 1% most of the time.

However, several things happened as the non-aggregate spatial parameter ($\phi_{z,1}$) increased, especially to 0.8. First, the proportion of times the aggregate residuals were white noise decreased when aggregating adjacent pairs of locations together ($q = 8$). Also occuring were steep decreases in the proportion of times the aggregate model produced better forecasts. This resulted in a larger increase in MSFE for using the aggregate model, on the order of 3 to over 5%. This is primarily due to the measure of misspecification error ($\mathbf{C}$) being a function of $\phi_{z,1}$. The more the non-aggregate locations are dependent on one another, the more evident the misspecification due to aggregation is.

Next we turn our attention to the parameter estimation in these cases. Table 2 provides the estimates of $\phi_{y,0}$, while Table 3 gives those of $\phi_{y,1}$. It should be noted that $\phi_{y,1}$ is not estimated in the case where $q = 1$ due to there being only one aggregate location and thus no spatial component. The primary phenomenon we see is that as the level of aggregation increases (more non-aggregate locations summed together), the more the aggregate autoregressive parameter increases and the aggregate spatial parameter decreases. Similar to Theorem 2, $\phi_{y,0}$ is greater than $\phi_{z,0}$ by a proportion of $\phi_{z,1}$, and $\phi_{y,1}$ is less than $\phi_{z,1}$ by a proportion of $\phi_{z,1}$. Thus, spatial aggregation transfers some of the spatial dependence into purely autoregressive dependence.

4. Conclusion

We have shown through simulation that in most cases of a realistic example of spatially aggregating a STAR(1_1) model, the aggregate also follows a STAR(1_1) model. This occurs even if the poolability condition does not hold exactly. In addition, the forecasts from the approximately poolable aggregate models are only slightly poorer on average than those from the non-aggregate model. The exception occurs for non-aggregate locations that have a high level of spatial dependence.

Also of interest was the discovery that the parameters of the aggregate STAR(1_1) model are nearly functions of the parameters in the non-aggregate model. For positive non-aggregate parameters, aggregation results in higher aggregate autoregressive parameters and lower aggregate spatial parameter. As a result, spatial aggregation can cause autoregressiveness in non-aggregate locations with little or no autoregressiveness. And for locations to display very little spatial correlation could be the result of very heavy spatial aggregation.

In future work, we will seek to broaden these results for more general setups of non-aggregate arrangement and form of aggregation. In addition it would be of benefit to extend these results to other forms of STARMA models. Lastly, we see potential in examining spatial aggregation among generalized STARMA (GSTARMA) models, developed by Borovkova et al. (2008), which have parameters that can vary by location. This could reduce the misspecification error caused by aggregation and the restricted nature of the weighting matrices.

References

Anselin, L. (1988). *Spatial econometrics: methods and models*, volume 4. Springer.

Arbia, G., Bee, M., and Espa, G. (2010). Aggregation of regional economic time series with different spatial correlation structures. *Geographical Analysis*, 43(1).

Borovkova, S., Lopuhaä, H. P., and Ruchjana, B. N. (2008). Consistency and asymptotic normality of least squares estimators in generalized star models. *Statistica Neerlandica*, 62(4):482–508.

Bun, M. J. (2004). Testing poolability in a system of dynamic regressions with nonspherical disturbances. *Empirical Economics*, 29(1):89–106.

Cliff, A. and Ord, J. K. (1975). Space-time modelling with an application to regional forecasting. *Transactions of the Institute of British Geographers*, pages 119–128.

Deutsch, S. J. and Pfeifer, P. E. (1981). Space-time arma modeling with contemporaneously correlated innovations. *Technometrics*, 23(4):401–409.

Giacomini, R. and Granger, C. W. (2004). Aggregation of space-time processes. *Journal of econometrics*, 118(1):7–26.

Griffith, D. A. (1988). *Advanced spatial statistics: special topics in the exploration of quantitative spatial data series*. Kluwer Academic Publishers.

Hendry, D. F. and Hubrich, K. (2006). Forecasting economic aggregates by disaggregates.

Hendry, D. F. and Hubrich, K. (2011). Combining disaggregate forecasts or combining disaggregate information to forecast an aggregate. *Journal of Business & Economic Statistics*, 29(2).

Kohn, R. (1982). When is an aggregate of a time series efficiently forecast by its past? *Journal of Econometrics*, 18(3):337–349.

Ljung, G. M. and Box, G. E. (1978). On a measure of lack of fit in time series models. *Biometrika*, 65(2):297–303.

Lutkepohl, H. (1984). Forecasting contemporaneously aggregated vector arma processes. *Journal of Business & Economic Statistics*, 2(3):201–214.

Lutkepohl, H. (1987). *Forecasting aggregated vector ARMA processes*, volume 284. Springer.

Lutkepohl, H. (2009). Forecasting aggregated time series variables: A survey.

Pfeifer, P. E. and Deutrch, S. J. (1980). A three-stage iterative procedure for space-time modeling. *Technometrics*, 22(1):35–47.

Pfeifer, P. E. and Jay Deutsch, S. (1980). Stationarity and invertibility regions for low order starma models: Stationarity and invertibility regions. *Communications in Statistics-Simulation and Computation*, 9(5):551–562.

Pino, F. A., Morettin, P. A., and Mentz, R. P. (1987). Modelling and forecasting linear combinations of time series. *International Statistical Review/Revue Internationale de Statistique*, pages 295–313.

Wei, W. W. and Abraham, B. (1981). Forecasting contemporal time series aggregates. *Communications in Statistics-Theory and Methods*, 10(13):1335–1344.

	A) % White Noise			B) % Agg. Lower			C) % Error Diff.		
	IND.	MOD.	HIGH	IND.	MOD.	HIGH	IND.	MOD.	HIGH
$(0.10, 0.10)$									
$q = 8$	85.6	86.0	86.0	36.8	32.8	48.4	0.10	0.12	0.11
$q = 4$	89.6	88.8	84.0	36.0	37.6	46.0	0.18	0.17	0.28
$q = 2$	88.8	86.0	93.6	36.4	37.6	44.4	0.45	0.34	0.39
$q = 1$	90.4	88.4	92.8	38.0	36.8	44.8	0.38	0.43	0.38
$(0.45, 0.10)$									
$q = 8$	83.6	85.6	84.4	33.2	36.0	46.4	0.11	0.14	0.11
$q = 4$	88.4	90.0	84.4	39.2	39.6	42.4	0.17	0.20	0.24
$q = 2$	89.2	84.8	94.4	36.4	35.2	42.4	0.50	0.41	0.41
$q = 1$	89.2	88.4	93.6	38.8	40.8	43.2	0.41	0.49	0.40
$(0.80, 0.10)$									
$q = 8$	84.0	82.4	85.2	27.6	34.4	41.2	0.26	0.20	0.15
$q = 4$	88.4	87.6	83.6	33.2	36.8	40.8	0.34	0.29	0.37
$q = 2$	88.8	86.0	94.0	35.6	35.6	39.6	0.74	0.62	0.73
$q = 1$	88.8	88.8	90.8	38.8	39.6	38.8	0.65	0.71	0.72
$(0.10, 0.45)$									
$q = 8$	76.4	64.0	75.2	6.0	14.4	36.4	1.39	0.85	0.18
$q = 4$	88.4	88.4	84.4	15.6	32.0	44.4	1.13	0.43	0.24
$q = 2$	89.6	84.8	93.6	20.4	35.2	42.0	1.45	0.57	0.40
$q = 1$	89.2	88.4	93.2	31.6	36.0	43.2	1.17	0.61	0.38
$(0.45, 0.45)$									
$q = 8$	69.6	47.2	67.2	4.4	13.6	35.6	1.95	1.30	0.21
$q = 4$	87.2	88.0	83.6	12.8	29.6	44.8	1.41	0.55	0.31
$q = 2$	89.6	84.4	94.4	18.8	33.2	41.2	1.85	0.80	0.65
$q = 1$	89.2	89.6	90.8	26.4	36.0	42.8	1.55	0.87	0.64
$(0.10, 0.80)$									
$q = 8$	37.2	5.2	31.2	0.0	2.8	24.0	5.61	3.22	0.48
$q = 4$	86.8	86.4	83.6	4.0	22.0	44.4	3.55	1.08	0.35
$q = 2$	89.6	82.8	94.4	6.4	28.4	41.6	3.99	1.21	0.68
$q = 1$	89.6	89.2	90.4	17.6	32.8	44.4	3.40	1.16	0.64

Table 1: Percentage of: A) Times Error in Aggregate Model is White Noise; B) Times Aggregate Model has Lower MSFE; C) Increase in MSFE in Aggregate Model

	A) Mean			B) Std. Dev.		
	IND.	MOD.	HIGH	IND.	MOD.	HIGH
$(0.10, 0.10)$						
$q = 8$	0.129	0.130	0.129	0.025	0.027	0.025
$q = 4$	0.164	0.168	0.165	0.038	0.035	0.035
$q = 2$	0.179	0.180	0.172	0.051	0.047	0.050
$q = 1$	0.194	0.195	0.191	0.068	0.069	0.069
$(0.45, 0.10)$						
$q = 8$	0.476	0.480	0.476	0.022	0.025	0.023
$q = 4$	0.510	0.516	0.511	0.032	0.032	0.030
$q = 2$	0.523	0.526	0.517	0.042	0.041	0.044
$q = 1$	0.537	0.541	0.533	0.056	0.059	0.059
$(0.80, 0.10)$						
$q = 8$	0.825	0.829	0.824	0.015	0.017	0.015
$q = 4$	0.856	0.863	0.858	0.020	0.020	0.020
$q = 2$	0.866	0.870	0.866	0.025	0.025	0.028
$q = 1$	0.880	0.882	0.882	0.034	0.036	0.037
$(0.10, 0.45)$						
$q = 8$	0.255	0.267	0.255	0.024	0.026	0.025
$q = 4$	0.412	0.434	0.413	0.033	0.033	0.032
$q = 2$	0.475	0.489	0.469	0.044	0.042	0.046
$q = 1$	0.536	0.546	0.533	0.055	0.059	0.059
$(0.45, 0.45)$						
$q = 8$	0.614	0.627	0.613	0.020	0.025	0.021
$q = 4$	0.765	0.783	0.767	0.023	0.024	0.024
$q = 2$	0.824	0.834	0.822	0.028	0.027	0.032
$q = 1$	0.881	0.885	0.882	0.034	0.035	0.037
$(0.10, 0.80)$						
$q = 8$	0.391	0.413	0.390	0.023	0.026	0.023
$q = 4$	0.670	0.700	0.672	0.026	0.027	0.026
$q = 2$	0.779	0.795	0.777	0.031	0.029	0.035
$q = 1$	0.880	0.886	0.882	0.034	0.035	0.037

Table 2: Estimates of Aggregate Autoregressive Parameter $\phi_{y,0}$: A) Mean; B) Standard Deviation

	A) Mean			B) Std. Dev.		
	IND.	MOD.	HIGH	IND.	MOD.	HIGH
$(0.10, 0.10)$						
$q = 8$	0.062	0.066	0.059	0.039	0.033	0.039
$q = 4$	0.029	0.025	0.029	0.049	0.040	0.051
$q = 2$	0.013	0.015	0.019	0.048	0.037	0.049
$(0.45, 0.10)$						
$q = 8$	0.061	0.064	0.059	0.034	0.029	0.036
$q = 4$	0.029	0.024	0.028	0.040	0.035	0.045
$q = 2$	0.012	0.013	0.016	0.040	0.032	0.041
$(0.80, 0.10)$						
$q = 8$	0.060	0.062	0.061	0.023	0.018	0.023
$q = 4$	0.028	0.023	0.029	0.026	0.022	0.025
$q = 2$	0.013	0.011	0.015	0.027	0.022	0.026
$(0.10, 0.45)$						
$q = 8$	0.280	0.292	0.278	0.035	0.031	0.037
$q = 4$	0.127	0.113	0.127	0.043	0.037	0.046
$q = 2$	0.060	0.055	0.064	0.042	0.033	0.043
$(0.45, 0.45)$						
$q = 8$	0.273	0.271	0.273	0.026	0.025	0.029
$q = 4$	0.121	0.107	0.120	0.030	0.026	0.032
$q = 2$	0.057	0.051	0.058	0.030	0.023	0.029
$(0.10, 0.80)$						
$q = 8$	0.494	0.485	0.493	0.028	0.018	0.031
$q = 4$	0.216	0.191	0.217	0.032	0.029	0.033
$q = 2$	0.100	0.091	0.103	0.033	0.025	0.032

Table 3: Estimates of Aggregate Spatial Parameter $\phi_{y,1}$: A) Mean; B) Standard Deviation