A Nonlinear Panel Data Model of Cross-sectional Dependence*

George Kapetanios
Queen Mary, University of London

James Mitchell
National Institute of Economic and Social Research, London

Yongcheol Shin
Leeds University Business School

November 22, 2010

Abstract

This paper proposes a new panel model of cross-sectional dependence. The model has a number of potential structural interpretations that relate to economic phenomena such as herding in financial markets. On an econometric level, it provides a flexible approach of modelling interactions across panel units and can generate endogenous cross-sectional dependence that can resemble the dependence that arises in a variety of existing models, such as factor or spatial models. We discuss the theoretical properties of the model and ways in which inference can be carried out. We supplement this analysis with a detailed Monte Carlo study and two empirical illustrations.

JEL Classification: C31, C33, C51, E31, G14.

Keywords: Nonlinear Panel Data Model, Clustering, Cross-section Dependence, Factor Models, Fixed Effects, Monte Carlo Simulations, Application to Stock Returns and Inflation Expectations

*We would like to thank seminar participants at Queen Mary, University of London, Essex University, City University, University of Piraeus, Yonsei University, conference delegates at the Econometric Society World Congress 2010 at Shanghai International Convention Center during 16-22 August, and especially Jinseo Cho, In Choi, Taewhan Kim, Jinwook Jeong, Myunghwan Seo for helpful comments and suggestions. Yongcheol Shin gratefully acknowledges partial financial support from the ESRC (grant No. RES-000-22-3161). The usual disclaimer applies.

1 Introduction

In many theoretical models economic agents learn from each other. Whether in herding models, where agents are assumed fully rational but have incomplete information sets (e.g., Banerjee (1992) and Bikhchandani, Hirshleifer, and Welch (1992)),[1] or in adaptive models where agents learn or form their expectations based on recent experience (see, e.g., Timmermann (1994) and Chevillon, Massmann, and Mavroeidis (2010)), agents are affected by past outcomes or the views of groups of other agents. Carroll (2003), for example, sets out a model whereby agents update their views probabilistically by looking at media reports, as opposed to forming full-information rational expectations. Similarly, cognitive psychology might be used to explain the contagion of views which leads to herd or imitating behaviour (see, e.g., Jegadeesh and Kim (2010)). See Akerlof and Shiller (2009) for a popular textbook discussion.

In this paper, motivated by this largely theoretical literature, we develop a general econometric modelling framework that incorporates herding effects and allows cross-sectional dependence, of many forms, to arise endogenously. In contrast, the popular factor models view cross-sectional dependence as an exogenous feature of the data. The proposal, discussion and econometric analysis of the proposed model, which is shown to nest many extant models as a special case, forms the main aim of this paper.

The model proposed in this paper is a nonlinear panel data model. Its distinguishing characteristic is the use of unit-specific aggregates of past values of variables relating to other units that are 'close' in some sense to a given unit, for the modelling of that unit. The nature of the model is dynamic in the sense that the past values of aggregates determine the present. We consider a number of nonlinear specifications for the construction of the unit specific aggregates, and our central specification is based on a threshold mechanism. The model nests a variety of dynamic panel data models, such as the standard panel data AR model. More interestingly, it is closely related to factor models that have received considerable attention in recent years following the works of Bai and Ng (2002), Stock and Watson (2002), and Bai (2003).

Our model provides a natural way in which many forms of cross-sectional dependence can arise in a large panel dataset comprised of variables of a similar nature that relate to different agents/units. The degree of cross-sectional dependence can vary, from a case where it is similar to standard factor models, for which the largest eigenvalue of the variance covariance matrix of the data tends to infinity at a rate N, where N is the number of cross-sectional units in the dataset, to the case of very weak or no factor structure where this eigenvalue is bounded as $N \to \infty$. Of course, all intermediate cases can arise as well. In this sense our work is closely

[1]Information-driven herding can sometimes be classified as "clustering" to differentiate it from herding due to extraneous incentive structures (e.g., Trueman (1994) and Hirshleifer and Teoh (2003)).

related to the works of Chudik and Pesaran (2010) and Chudik, Pesaran, and Tosetti (2009). These papers discuss the concepts of weak and strong cross-sectional dependence based on the characteristics of the variance-covariance matrix of the data and are dynamic in nature, being instances of large dimensional VAR models. Our work can be viewed as a particular instance of a large dimensional VAR but for the fact that our model is intrinsically nonlinear in nature. More importantly, our work is possibly the first specific instance of a model that falls within the remit of the general class of models discussed in Chudik and Pesaran (2010), has a flexible way of allowing many forms and degrees of cross-sectional dependence and has a clear structural interpretation that relates to the intuitive herding ideas discussed above.

Our work has precedents in the system engineering literature. However, all the work in that literature relaters to deterministic models whose limit behaviour is a fixed point that represents clustering. A discussion of the asymptotic behaviour of the deterministic version of our basic model can be found in Blondel, Hendrickx, and Tsitsiklis (2009) following on from the work of Krause (1997). Another literature that is closely related to our work is the 'similarity' literature as exemplified by Gilboa, Lieberman, and Schmeidler (2006) and references therein. This work relates to univariate processes. It suggests that forecasting for y_t, at time T, can be based on a model which places heavier weights on those past observations of y_t, for which a given vector of variables, x_t, is close to x_T with respect to some metric. In other words, observations y_t, $t \leq T$, for which $||x_t - x_T||$ is small, for some metric $||.||$, have a larger weight for constructing forecasts of y_{T+1} at time T. Gilboa, Lieberman, and Schmeidler (2006) provide powerful theoretical economic justifications for this approach. Our work can be thought of as an extension of this analysis to a multi-agent panel framework, where similarity between agents takes the place of similarity between circumstances.

We provide a comprehensive analysis of the stochastic version of the model, while allowing both for threshold but also smooth transition type nonlinearities. Further, we discuss estimation of the model and analyse the asymptotic properties of the estimators. We propose a large number of extensions to the basic model that cover many interesting cases. These include the combination of our 'herding' mechanism with more traditional forms of factor modelling that allow for exogenous forms of cross-sectional dependence. As another example of an extension, a more realistic modelling of the very complicated behaviour of and interactions among agents can be analysed in a flexible manner by employing neural network type specifications in conjunction with partial cross-sectional averaging. This extension could shed further lights on identifying the possibly asymmetric impacts of the differences of opinions on stock prices and volumes (see, e.g., Banerjee, Kaniel, and Kremer (2009) and Banerjee and Kremer (2010)).

A very interesting aspect of our work arises out of our analysis of how the new model, being a generalised autoregressive panel data model, relates to the simple panel data AR model. Interestingly, we find that the standard 'Nickel' bias (c.f., Nickell (1981)) that arises in the simple panel data AR model and leads to the need for the IV (more generally GMM) estimation, does not arise in a number of our specifications, but does in others.

The structure of the paper is as follows: Section 2 presents the basic specification of the model and discusses in detail its theoretical properties. Section 3 presents a number extensions and discusses their properties. Section 4 discusses the issue of how to test for the presence of nonlinearity in the data. Section 5 presents extensive Monte Carlo simulation evidence. Section 6 provides two empirical illustrations for analysing nonlinearity and cross-section dependence of stock returns and inflation expectations, which clearly demonstrate the usefulness of our proposed models. 7 concludes. All proofs are relegated to an Appendix.

2 The Theoretical Model

We propose a model, which can be given a behavioural interpretation, based on the commonplace idea that agents consider the views or behaviour of those around them and aggregate them in some way in order to decide on their own expectations or behaviour. This interaction or mimicking may be explicit, in the sense that agents know what the other agents experienced or expect, or could be implicit, in the sense that groups of agents happen to behave similarly even though they do not interact formally. This might be because they are subject to the same environment and/or have similar information sets when forming expectations. To formalise this idea, we propose an explicit dynamic panel model of a multitude of agents. Let $x_{i,t}$ denote the value of the variable of interest, such as the agent's income or the agent's view of the future value of some macroeconomic variable, at time t, for agent i. We assume a sample of T observations for each of N agents. Then, we specify that

$$x_{i,t} = \frac{\rho}{m_{i,t}} \sum_{j=1}^{N} \mathcal{I}\left(|x_{i,t-1} - x_{j,t-1}| \leq r\right) x_{j,t-1} + \epsilon_{i,t}, \quad t = 2, ..., T, \quad i = 1, ..., N, \tag{1}$$

where

$$m_{i,t} = \sum_{j=1}^{N} \mathcal{I}\left(|x_{i,t-1} - x_{j,t-1}| \leq r\right),$$

$\{\epsilon_{i,t}\}_{t=1}^{T}$ is an error process whose properties will be further discussed below, $\mathcal{I}(.)$ is the indicator function and $-1 < \rho < 1$. Verbally, the above model states that $x_{i,t}$ is influenced by the cross-sectional average of a selection of past x_j and in particular that the relevant x_j are those that lie closest to $x_{i,t-1}$. This formalises the intuitive idea that people are affected more by those with whom they share common views or behaviour. The model may be equally viewed as a descriptive model of agents' behaviour, reflecting the fact that 'similar' agents are affected by 'similar' effects, or as a structural model of agents' views whereby agents use the past views of other agents, similar to them in some respect, to form their own views. The interactive term in (1) may then be thought to capture the (cross-sectional) local average or common component of their views. This idea of commonality has various clear, motivating, concrete examples in a variety of social science disciplines such as psychology and politics. In economics and finance, the herding could be rational (imitative herding: see Devenow and Welch (1996)) or irrational.

A deterministic form of the above model has been analysed previously in the mathematical and system engineering literature. In particular, Blondel, Hendrickx, and Tsitsiklis (2009) have analysed a continuous form of the restricted version of (1) given by

$$x_{i,t} = \frac{1}{m_{i,t}} \sum_{j=1}^{N} \mathcal{I}\left(|x_{i,t-1} - x_{j,t-1}| \leq 1\right) x_{j,t-1}, \quad t = 2, ..., T, \quad i = 1, ..., N. \tag{2}$$

where $m_{i,t} = \sum_{j=1}^{N} \mathcal{I}\left(|x_{i,t-1} - x_{j,t-1}| \leq 1\right)$. To the best of our knowledge, we are the first to introduce a stochastic term to this type of model and to allow for an unknown value of the threshold parameter.

(1) bears considerable resemblance to threshold autoregressive (TAR) models analysed in the time series literature. However, unlike straightforward extensions of such models to a panel setting whereby individual units/agents would not have interactions through the nonlinear specification, the nonlinearity in (1) is inherently cross-sectional in nature and provides for the development of a dynamic network effect. In deterministic contexts this has been shown to generate interesting behaviour, such as clustering.

2.1 Clustering

To appreciate more concretely the dynamics of the model we report various graphical results. We start by reporting the dynamic behaviour of the deterministic model (i.e. setting $\epsilon_{i,t} = 0$). In particular we set $N = 100$, $T = 20$. We set the initial conditions to $x_{i,0} \sim N(0, 25)$ and report the evolution of the system for $\rho = 1$, $r = 0.5$ and 3, in Figure 1. As we see, the system settles quickly to a steady state with a number of clusters. The number of clusters declines with the size of the threshold parameter, as one would intuitively expect. Obviously, for a large value of r, only one cluster will arise.

Of course, the dynamic behaviour of the stochastic model can be expected to be quite different. To explore this, we present some realisations of the stochastic system. We set $N = 100$, $T = 500$, and the initial conditions as before. For the rest of the parameters, we set $r = 0.5$, $\rho = 0.999$, and $\epsilon_{i,t} \sim N(0, 0.1)$. As we will discuss below, the model is stationary for $|\rho| < 1$, and the behaviour of a stationary model is of particular interest. But nonstationarity is of interest, too, and has been extensively explored in the literature dealing with factor models. The most interesting behaviour of the model can be obtained when ρ is high enough for the model to be quite persistent. We report two realisations of this model in Figure 2. The first realisation shows emerging cluster structures in the first 100 observations. Then, there are clearly two clusters that persist throughout the rest of the sample. A number of units are outlying and do not join any cluster for the whole sample. The second realisation has one dominant cluster. There is a second cluster which starts at the beginning of the sample and fizzles out by observation 250. At that point a new cluster emerges and by the end of the sample becomes as dominant as the original major cluster.

Clearly this model can model flexibly all sorts of clustering behaviour. It is tempting to attempt to characterise the behaviour of the model as a function of the parameters and it is clear that for persistent ρ, the interplay of r and the variance of $\epsilon_{i,t}$ is crucial. For instance, a small variance for $\epsilon_{i,t}$ relative to r implies that units do not escape clusters easily. Similarly, ceteris paribus, a larger r leads to fewer clusters and dynamically to faster consolidation towards clusters. This needs to be tempered with the finding, discussed in detail later, than when the value of r tends to infinity the model has a smaller degree of cross-sectional dependence. So, overall it seems that the model can behave in distinct ways depending sensitively on all its parameters, including higher moments of $\epsilon_{i,t}$, as we discuss below.

Next, we wish to allow for fat tails in the distribution of $\epsilon_{i,t}$. Therefore, we set $\epsilon_{i,t} \sim t_3$, and subsequently normalise $\epsilon_{i,t}$ to have variance equal to 0.1. We report a realisation of this model in Figure 3. Here, it is clear that more clusters arise. There is cluster consolidation but at the same time cluster bifurcation (see the cluster made up of units with high values that bifurcates around observation 400 only to re-emerge as a single cluster by the end of the sample). Overall, it is clear that the new model can generate complex behaviour across units.

2.2 Special cases

It is interesting to note the nature of restricted versions of the above model, obtained by taking extreme values of the threshold parameter. By setting $r = 0$, we obtain a simple panel autoregressive model of the form

$$x_{i,t} = \rho x_{i,t-1} + \epsilon_{i,t} \tag{3}$$

On the other hand letting $r \to \infty$, we obtain the model

$$x_{i,t} = \frac{\rho}{N} \sum_{j=1}^{N} x_{j,t-1} + \epsilon_{i,t} \tag{4}$$

where past cross-sectional averages of opinions inform, in similar fashions, current opinions. Recently, the use of cross-sectional averages has been advocated by Pesaran (2006), Chudik and Pesaran (2010) and Chudik, Pesaran, and Tosetti (2009) as a means of modelling cross-sectional dependence in the form of unobserved factors. However, unlike these models where the use of cross-sectional averages is an approximation to the unknown model, in our case this is a limiting case of a structural nonlinear model.

A graphical comparison of these restricted versions of the nonlinear model is also instructive. In Figure 4, we report comparable realisations to those in Figure 1 but setting $r = 0$ in the upper panel and $r = \infty$ in the lower panel. These are, of course, just single realisations; but repeated realisations suggest a very similar picture. While the upper panel depicts independent and very persistent series evolving with little regard to other series in the panel, the lower panel depicts a closely linked set of series behaving similarly. It is interesting to note that this similarity, reminiscent of factor structures, can be proven to arise only for finite N when $|\rho| < 1$, as we will discuss in more detail below. Neither of these pictures compares in terms of complexity and flexibility to the realisations of the nonlinear model seen in Figures 2-3. It is clear that neither of these two restricted versions of the model can accommodate clustering or evolving herding.

It is important to investigate the properties of our model. A number of results, stated and proved in the appendix, provide help in this respect. Intuitively, as we show in Lemma 1, (1) is geometrically ergodic, and therefore asymptotically stationary, if $|\rho| < 1$. This allows for the analysis of estimators along traditional lines, as discussed below.

2.3 Cross-sectional dependence and factor models

It is of interest to examine the cross-sectional dependence properties of the model. This is slightly complicated by the need to define cross-sectional dependence in our context. We choose to follow an approach which is used in the analysis of factor models. In the factor literature, the behaviour of the covariance matrix of $x_t = (x_{1,t}, ..., x_{N,t})'$, is considered. Factor models have the property that both the maximum eigenvalue and the row/column sum norm of the covariance matrix tend to infinity at rate N, as $N \to \infty$. In contrast, for other models of cross-sectional dependence such as, for example, spatial AR or MA models, these quantities are bounded, implying that they exhibit much lower degrees of cross-sectional dependence than factor models.[2] It is useful to see where our model fits in this nomenclature. Lemma 4 shows that the column sum norm of the variance covariance matrix of x_t when x_t follows (1) is $O(N)$. Thus, the model is much more similar to factor models than spatial AR or MA models. Interestingly, as we will see in the next section that discusses extensions to the basic model (1), there are versions of (1) that resemble spatial models, more than factor models. Another very interesting finding is that (4) implies a variance covariance matrix for x_t with a column sum norm that is $O(1)$. This is surprising, given the similarity that cross-sectional average schemes have with factor models as detailed in Pesaran (2006). However, this result and the analysis of Pesaran (2006) are not directly comparable. Pesaran (2006) assumes the prior existence of factors and uses cross-sectional averages to approximate the existing factors.

[2] A useful discussion of the various concepts of cross-sectional can be found in Chudik and Pesaran (2010)

These pre-existing exogenous factors generate high cross-sectional dependence and herding. In our case no exogenous factors exist and the cross-sectional average is a primitive term that exists in the structure of the model. Our surprising result is proven in Lemma 3.[3]

Given the above, it is of interest to examine the analogy with factor models in more detail. We do this by simulating data using (1) and the parametrisation used to construct the realisations in Figure 2. Using the simulated dataset we then extract factor estimates using principal components. We extract 8 principal components and subsequently examine the proportion of the variance of the dataset explained by these principal components. Our previous pictorial analysis suggests that factor like behaviour emerges in the form of clusters of series moving together. The first column of Table 1 presents the average cumulative proportion of the dataset variance explained by successive principal components, over 100 replications. As we can see there is behaviour reminiscent of factor analysis. The first factor explains about 40% of the total dataset variance rising to about 77% when all 8 factors are considered.

For comparability, we also consider simulations from the same model but setting $r = \infty$. Results are reported in the second column of Table 1. As we see, while the first factor explains roughly the same proportion of the variance in the two parametrisations, the rest of the factors explain little further. This is reasonable. In this case there is only one cluster arising around the cross-sectional mean. As we noted above, there is a crucial difference between (1) and (4). This relates to the fact that while the column sum norm of x_t for (1) is $O(N)$, it is $O(1)$ for (4). This result is asymptotic with respect to N and as noted in footnote 3, the distinction can be difficult to discern for values of ρ close to 1. As a result, we consider a further simulation along the same lines but setting higher values for N ($N = 100, 200, 400, 800, 1000$ and 1500) and a lower value for ρ ($\rho = 0.8$). Results on the average cumulative proportion of the dataset variance explained by successive principal components, over 100 replications, are reported in Tables 2 and 3. It is clear that data from (1) are more cross-sectionally dependent than data from (4). More pertinently, while it is clear that as N increases principal components can explain a decreasing proportion of the data variance for (4), the proportion remains constant for (1).

It is important to restate here one crucial difference between our model and a factor model. Our model has a clear parametric structure and its properties as an approximating mechanism for generic cross-sectional dependence is, and, to some extent, should be, unclear and limited. On the other hand, a factor model can lay claim to some generality, in the following sense. Once a dataset has pronounced cross-sectional dependence exhibited by, say, explod-

[3]It is interesting to note that further interesting interactions arise if we let $\rho = 1$. This unit root behaviour counteracts the tendency of the cross-sectional average to disappear asymptotically as $N \to \infty$. Then, the behaviour of both the variances and the covariances of x_t as both N and $T \to \infty$, depends on the limit of $\frac{T}{N}$. For example, as long as $\frac{T}{N}$ remains bounded so do the variances of x_t, despite the unit root structure of the model. We feel that a detailed investigation of this possibility is beyond the scope of the present paper.

ing eigenvalues or column sum norms, associated with its covariance matrix, then a factor model should have good approximation ability irrespective of the structural form giving rise to the cross-sectional dependence. In a similar vein, in a strongly cross-sectionally dependent dataset, principal components are able nonparametrically to construct linear combinations of the variables that can capture this cross-sectional dependence, irrespective of its origin, as we have seen above. Of course, since our model nests (4), it is reasonable to expect that it can approximate a factor model by allowing $r \to \infty$, in a similar manner to that underlying the analysis of Pesaran (2006). Finally, it is worth noting that the factor model, unlike the nonlinear model (or its relevant extensions in Section 3), cannot accommodate the case when the cross-sectional dependence is not strong as in the case of spatial models.

2.4 Estimation

In this section we explore estimation of the nonlinear model in (1). We consider the standard estimation procedure for a threshold model, whereby a grid of values for r is constructed. Then for all values on that grid the model is estimated by least squares to obtain estimates of the autoregressive parameter, ρ. More specifically, denoting $\tilde{x}_{i,t} = \frac{1}{m_{i,t}} \sum_{j=1}^{N} \mathcal{I}\left(|x_{i,t-1} - x_{j,t-1}| \leq r\right) x_{j,t-1}$, $\tilde{x}_i = (\tilde{x}_{i,1}, ..., \tilde{x}_{i,T-1})'$, $\tilde{x} = (\tilde{x}_1', ..., \tilde{x}_N')'$, $x_i = (x_{i,2}, ..., x_T)'$ and $x = (x_1', ..., x_N')'$, x is regressed on $\tilde{x}$ using OLS to give an estimate for ρ, for a given value of r in the grid. The value of r that minimises the sum of squared residuals, $\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{\epsilon}_{i,t}^2(\rho, r)$, where

$$\hat{\epsilon}_{i,t}(\rho, r) = x_{i,t} - \frac{\rho}{m_{i,t}} \sum_{j=1}^{N} \mathcal{I}\left(|x_{i,t-1} - x_{j,t-1}| \leq r\right) x_{j,t-1}$$

is the estimator of r. We denote the least squares estimator of (ρ, r) by $(\hat{\rho}, \hat{r})$. We make the following assumption about the error term, $\epsilon_{i,t}$.

Assumption 1 *$\epsilon_{i,t}$ is i.i.d. across t and independent across i. $E(\epsilon_{i,t}^2) = \sigma_{\epsilon_i}^2$. $E(\epsilon_{i,t}^4) < \infty$. For all i, the density of $\epsilon_{i,t}$ is bounded and positive over all compact subsets of $\mathbb{R}$.*

Then, we have the following theorems:

Theorem 1 *Let Assumption 1 hold for $\epsilon_{i,t}$ in (1). Then, as long as $|\rho| < 1$, the least squares estimator of (ρ, r) is consistent as $N, T \to \infty$.*

Theorem 2 *Let Assumption 1 hold for $\epsilon_{i,t}$ in (1). Let (ρ^0, r^0) denote the true value of (ρ, r). Then, as long as $|\rho| < 1$, $NT(\hat{r} - r^0) = O_p(1)$. Further, as long as $|\rho| < 1$, $(NT)^{1/2}(\hat{\rho} - \rho^0)$ has the same asymptotic distribution as if r^0 was known.*

These theorems are intuitive, as they accord with the work and theoretical analysis of Chan (1993) who was the first to analyse, theoretically, the estimator for the univariate threshold autoregressive model. There exist a number of possible theoretical extensions of this estimation problem. One obvious one relates to the fact that the asymptotic distribution of $NT(\hat{r} - r^0)$ is non-normal and depends on unknown parameters, as discussed in Chan (1993). The work of Hansen (2000) is of great use here, since by assuming that the model is linear asymptotically, a tractable distributional theory can be obtained for $\hat{r}$. We feel that it is perhaps more appropriate to allow for the nonlinearity to persist asymptotically and, therefore, we do not pursue further this interesting avenue of research.

2.5 Unbalanced panels

The model in (1) can be adjusted to allow for unbalanced panels. In this case (1) takes the form

$$x_{i,t} = \rho \tilde{x}_{i,t}^{up} + \epsilon_{i,t}, \quad t = 2, ..., T, \quad i = 1, ..., N_t, \tag{5}$$

as long as both $x_{i,t}$ and $\tilde{x}_{i,t}^{up}$ are observable, where N_t is the number of observable pairs, $\left(x_{i,t}, \tilde{x}_{i,t}^{up}\right)$, at time t. The definition of $\tilde{x}_{i,t}^{up}$ depends on the application at hand. An obvious definition is

$$\tilde{x}_{i,t}^{up} = \frac{\rho}{m_{i,t}} \sum_{j=1}^{N_{t-1}} \mathcal{I}\left(|x_{i,t-1} - x_{j,t-1}| \leq r\right) x_{j,t-1} \tag{6}$$

where $m_{i,t} = \sum_{j=1}^{N_{t-1}} \mathcal{I}\left(|x_{i,t-1} - x_{j,t-1}| \leq r\right)$ and $(x_{i,t}, x_{i,t-1})$ is observable. Alternative specifications can be used to increase the number of available observations. For example, if $x_{i,t-1}$ is not observed, the latest available observation for the i-th unit prior to time t could be used. More specifically, letting $s_{i,t}$ denote the latest time period, prior to t, in which x is observable for unit i, we can define $\tilde{x}_{i,t}^{up}$ as either

$$\tilde{x}_{i,t}^{up} = \frac{\rho}{m_{i,t}} \sum_{j=1}^{N_{t-1}} \mathcal{I}\left(\left|x_{i,s_{i,t}} - x_{j,t-1}\right| \leq r\right) x_{j,t-1} \tag{7}$$

or

$$\tilde{x}_{i,t}^{up} = \frac{\rho}{m_{i,t}} \sum_{j=1}^{N_{s_{i,t}}} \mathcal{I}\left(\left|x_{i,s_{i,t}} - x_{j,s_{i,t}}\right| \leq r\right) x_{j,s_{i,t}} \tag{8}$$

where

$$m_{i,t} = \sum_{j=1}^{N_{t-1}} \mathcal{I}\left(\left|x_{i,s_{i,t}} - x_{j,t-1}\right| \leq r\right)$$

and

$$m_{i,t} = \sum_{j=1}^{N_{s_{i,t}}} \mathcal{I}\left(\left|x_{i,s_{i,t}} - x_{j,s_{i,t}}\right| \leq r\right),$$

respectively. The specifications in (7) and (8) allow for a larger set of available observations to be used than in (6). Estimation of this model can then be carried out similarly to the case where the number of cross-sectional units is fixed over time. In this case, the effective number of observations is equal to the number of observable pairs of $\left(x_{i,t}, \tilde{x}_{i,t}^{up}\right)$ over i and t, rather than NT, and the statements of Theorems 1 and 2 need to be amended accordingly.

Model (1) can be extended in a large variety of ways. We explore a number in the next Section.

3 Extensions

The model given in (1), while interesting from the perspective of analysing cross-sectional dependence or studying phenomena, such as herding, in an empirical context is quite restrictive in a number of senses. This section provides some extensions that alleviate this. Given that our benchmark model is a panel model it is reasonable to include fixed effects. For such an extension, the basic model becomes

$$x_{i,t} = \nu_i + \frac{\rho}{m_{i,t}} \sum_{j=1}^{N} \mathcal{I}\left(|x_{i,t-1} - x_{j,t-1}| \leq r\right) x_{j,t-1} + \epsilon_{i,t} \tag{9}$$

where $\nu_i \sim i.i.d.(0, \sigma_\nu)$. Of course, more general versions of the above model can be accommodated, such as

$$x_{i,t} = \nu_i \zeta_t + \frac{\rho}{m_{i,t}} \sum_{j=1}^{N} \mathcal{I}\left(|x_{i,t-1} - x_{j,t-1}| \leq r\right) x_{j,t-1} + \epsilon_{i,t} \tag{10}$$

for an $r \times 1$ vector of observable variables, ζ_t.

We now examine the properties of the least squares estimator for (9). As is well known, the presence of ν_i induces endogeneity in standard panel AR models, leading to biased estimation of the autoregressive parameter for finite T, when standard panel least squares estimators, such as the within group estimator, are used. It is easiest to see the problem for standard AR models, and its relation to our model, by noting that the endogeneity arises because unbiasedness, for least squares estimators, requires that

$$E\left(x_{i,t-1}\left(\epsilon_{i,t} - \frac{1}{T}\sum_{t=1}^{T}\epsilon_{i,t}\right)\right) = 0 \tag{11}$$

Obviously, the expectation in (11) is not zero but $O\left(\frac{1}{T}\right)$. One would expect a similar problem to arise for (9). However, surprisingly, this is not the case. As is shown in Lemma 11 in the Appendix

$$E\left(\left(\frac{1}{m_{i,t}} \sum_{j=1}^{N} \mathcal{I}\left(|x_{i,t-1} - x_{j,t-1}| \leq r\right) x_{j,t-1}\right)\left(\epsilon_{i,t} - \frac{1}{T}\sum_{t=1}^{T}\epsilon_{i,t}\right)\right) = O\left(\frac{1}{NT}\right) \tag{12}$$

which implies that Theorems 1 and 2 hold for (9). As a result the standard within group estimator can be used for (9), thus removing the need for less efficient GMM estimation as is usually the case.

The restriction that (1) has a single lag can be relaxed to allow for p lags so that

$$x_{i,t} = \sum_{s=1}^{p} \left[\frac{\rho_s}{m_{i,t,s}} \sum_{j=1}^{N} \mathcal{I}\left(|x_{i,t-s} - x_{j,t-s}| \leq r\right) x_{j,t-s} \right] + \epsilon_{i,t} \tag{13}$$

where $m_{i,t,s} = \sum_{j=1}^{N} \mathcal{I}\left(|x_{i,t-s} - x_{j,t-s}| \leq r\right)$. Alternatively, we can introduce more regimes and consider the model given by

$$x_{i,t} = \sum_{s=1}^{q} \left[\frac{\rho_s}{m_{i,t,s}} \sum_{j=1}^{N} \mathcal{I}\left(r_s \leq |x_{i,t-1} - x_{j,t-1}| < r_{s+1}\right) x_{j,t-s} \right] + \epsilon_{i,t} \tag{14}$$

where $m_{i,t,s} = \sum_{j=1}^{N} \mathcal{I}\left(r_s \leq |x_{i,t-1} - x_{j,t-1}| < r_{s+1}\right)$. Both (13) and (14) can be estimated similarly to (1). However, sufficient conditions for their geometric ergodicity are different to those for (1), and are given in Lemmas 13 and 14, respectively.

As we noted in the introduction, (1) has a structural interpretation, whereby agents collect information about other agents' views and behaviour and use them to construct their own. But it can be reasonably argued that this information gathering has costs. As a result, and as $N \to \infty$, it may be reasonable to suppose that not all units close to i will be used when constructing cross-sectional averages. This idea of costs to information gathering can be formalised, so that the model may be modified to

$$x_{i,t} = \frac{\rho}{\tilde{m}_{i,t}} \sum_{j=1}^{\tilde{m}_{i,t}} x_{j,t-1}^{(i)} + \epsilon_{i,t} \tag{15}$$

where $x_{t-1}^{(i)} = (x_{1,t-1}^{(i)}, ..., x_{N,t-1}^{(i)})'$, $x_{t-1}^{(i)} = \texttt{sort}(x_{t-1}, (|x_{1,t-1} - x_{i,t-1}|, ..., |x_{N,t-1} - x_{i,t-1}|)')$,

$\texttt{sort}(a, b)$ sorts the vector a in the same order as the ascending order sort of the vector b and $\tilde{m}_{i,t} = \min(m_{i,t}, m)$, for some constant, m, possibly depending on T or N. This version of the model places the restriction that only the first m units closest to i at time $t-1$, enter the cross-sectional average, at most. This model has quite distinct properties. It is still geometrically ergodic if $|\rho| < 1$, as proven by Lemma 6. However, its cross-sectional dependence properties are different since it is much closer to a standard panel AR model. As proven in Lemma 7, the column sum norm of the covariance matrix of x_t, when x_t follows (15), is bounded as $N \to \infty$, as long as $m\rho < 1$. As a result, the extent of cross-sectional dependence for this variant of the model is much smaller. This model introduces another parameter that needs to be estimated: m. The estimation of m can be carried out similarly to the estimation of r, by constructing a two dimensional grid of values for (r, m) and then choosing the combination that minimises the sum of squared residuals. We have the following consistency result for this model, proven in the Appendix.

Theorem 3 *Let Assumption 1 hold for $\epsilon_{i,t}$ in (15). Then, as long as $|\rho| < 1$, the least squares estimator of (ρ, r, m) is consistent for finite N and as $T \to \infty$.*

Another implication of the reduced extent of cross-sectional dependence for this model relates to the extension of this model that allows for fixed effects. As we discussed above, (9) can be estimated using the within group estimator without suffering biases due to the presence of lagged endogenous variables. This is possible because the lagged term in (9) is sufficiently diluted by the presence of lagged variables belonging to other cross-sectional units to allow for (12) to hold. This is not the case for (15), as shown in Lemma 12, in the Appendix. As a result, estimation of this model using the within estimator is biased, possibly severely so, for finite T. The standard solution is to estimate by GMM. In our case the moment conditions take the form:

$$E\left(x_{i,t-s}\left(\Delta x_{i,t} - \rho\Delta\left(\frac{1}{\tilde{m}_{i,t}}\sum_{j=1}^{\tilde{m}_{i,t}} x_{j,t-1}^{(i)}\right)\right)\right) = 0, \quad t = 3, ..., T, \quad s = 1, .., t-2.$$

This is a set of $T(T-1)/2$ conditions that could, in principle, be used to estimate efficiently the model. However, there are problems with this approach. As noted in Caner and Hansen (2004), GMM estimation of threshold models is inconsistent if the variable used in the indicator function is endogenous, as it is in our case. Recently, Kourtellos, Tan, and Stengos (2008) have suggested a method for estimating, with GMM, threshold models with endogenous switches. But, the derivation of distributional results is not clear and so the applicability of the method is unclear too. Of course, the 'within' estimator is consistent as $T \to \infty$, and so can be used for long panels.

Up until now we have considered only threshold mechanisms for constructing the unit-specific cross-sectional averages. This need not be the case. In particular, we can envisage models of the form

$$x_{i,t} = \rho \sum_{j=1}^{N} \frac{w(|x_{i,t-1} - x_{j,t-1}|\,; \gamma) x_{j,t-1}}{\sum_{j=1}^{N} w(|x_{i,t-1} - x_{j,t-1}|\,; \gamma)} + \epsilon_{i,t} \tag{16}$$

where $w(x;\gamma)$ is a positive twice differentiable integrable function such as, e.g., the exponential function $\exp(-\gamma x^2)$ or the normal cdf, $\Phi(x)$. By now, the properties of this model should be reasonably clear. Lemma 8 shows that the model is geometrically ergodic if $|\rho| < 1$ and similarly to model (1), the column sum norm of the covariance matrix of x_t, when x_t follows (16) is $O(N)$, as shown in Lemma 9. The model in its simple form given by (16) can be estimated by nonlinear least squares; and we have the following Theorem concerning the asymptotic properties of this estimator.

Theorem 4 *Let Assumption 1 hold for $\epsilon_{i,t}$ in (16). Then, as long as $|\rho| < 1$, the nonlinear least squares estimator of (ρ, γ) is $(NT)^{1/2}$-consistent and asymptotically normal as $N, T \to \infty$.*

Similarly to Lemma 11, it can also be shown that

$$E\left(\left(\sum_{j=1}^{N} \frac{w(|x_{i,t-1} - x_{j,t-1}|\,; \gamma) x_{j,t-1}}{\sum_{j=1}^{N} w(|x_{i,t-1} - x_{j,t-1}|\,; \gamma)}\right)\left(\epsilon_{i,t} - \frac{1}{T}\sum_{t=1}^{T} \epsilon_{i,t}\right)\right) = O\left(\frac{1}{NT}\right) \tag{17}$$

which implies that a 'within' estimator is valid for estimating (16), when fixed effects are incorporated in (16). The above model can be refined to allow for costly information gathering. In particular, the following model has the required feature in a smooth transition framework

$$x_{i,t} = \rho \sum_{j=1}^{m} \frac{w\left(\left|x_{i,t-1} - x_{j,t-1}^{(i)}\right|; \gamma\right) x_{j,t-1}^{(i)}}{\sum_{j=1}^{m} w\left(\left|x_{i,t-1} - x_{j,t-1}^{(i)}\right|; \gamma\right)} + \epsilon_{i,t} \tag{18}$$

Again, similarly to Lemma 12, it can easily be seen that

$$E\left(\left(\sum_{j=1}^{m} \frac{w\left(\left|x_{i,t-1} - x_{j,t-1}^{(i)}\right|; \gamma\right) x_{j,t-1}^{(i)}}{\sum_{j=1}^{m} w\left(\left|x_{i,t-1} - x_{j,t-1}^{(i)}\right|; \gamma\right)}\right)\left(\epsilon_{i,t} - \frac{1}{T}\sum_{t=1}^{T} \epsilon_{i,t}\right)\right) = O\left(\frac{1}{T}\right) \tag{19}$$

implying that, once fixed effects are introduced, the within estimator is biased for finite T. Then, GMM based on the set of $T(T-1)/2$ conditions given by

$$E\left(x_{i,t-s}\left(\Delta x_{i,t} - \rho \Delta\left(\sum_{j=1}^{m} \frac{w\left(\left|x_{i,t-1} - x_{j,t-1}^{(i)}\right|; \gamma\right) x_{j,t-1}^{(i)}}{\sum_{j=1}^{m} w\left(\left|x_{i,t-1} - x_{j,t-1}^{(i)}\right|; \gamma\right)}\right)\right)\right) = 0, \quad t = 3, ..., T, \quad s = 1, .., t-2.$$

provides a consistent and asymptotically normal estimator for (ρ, γ) and a consistent estimator for m, through a grid search over possible values of m.

Another obvious extension to the set of models we have been developing is to introduce other variables to the model either linearly as in

$$x_{i,t} = \frac{\rho}{m_{i,t}} \sum_{j=1}^{N} \mathcal{I}\left(|x_{i,t-1} - x_{j,t-1}| \leq r\right) x_{j,t-1} + \beta z_{i,t} + \epsilon_{i,t} \tag{20}$$

or nonlinearly as in, e.g.,

$$x_{i,t} = \frac{\rho}{m_{i,t}} \sum_{j=1}^{N} \mathcal{I}\left(|x_{i,t-1} - x_{j,t-1}| \leq r\right) x_{j,t-1} + \frac{\beta}{m_{i,t}} \sum_{j=1}^{N} \mathcal{I}\left(|x_{i,t-1} - x_{j,t-1}| \leq r\right) z_{j,t-1} + \epsilon_{i,t} \tag{21}$$

or to introduce other switch variables, giving rise to a model of the form

$$x_{i,t} = \frac{\rho}{m_{i,t}} \sum_{j=1}^{N} \mathcal{I}\left(|x_{i,t-1} - x_{j,t-1}| \leq r_1\right) x_{j,t-1} + \frac{\beta}{m_{z,i,t}} \sum_{j=1}^{N} \mathcal{I}\left(|z_{i,t-1} - z_{j,t-1}| \leq r_2\right) x_{j,t-1} + \epsilon_{i,t} \tag{22}$$

where $m_{z,i,t} = \sum_{j=1}^{N} \mathcal{I}\left(|z_{i,t-1} - z_{j,t-1}| \leq r_2\right)$. It is also clear from the work of Kapetanios (2001) that information criteria can be used to choose the switch variables. The theoretical properties of the models in (20)-(22) should be obvious from the preceding analysis. For example, geometric ergodicity of (22) holds if $|\rho + \beta| < 1$. Another interesting point is that if costly information gathering is combined with a model with switches (such as, e.g., a model of the form (22) with only the second nonlinear term included, assuming that $z_{i,t}$, is strictly exogenous), then this model can be estimated via GMM using the following set of GMM conditions

$$E\left(x_{i,t-s}\left(\Delta x_{i,t} - \beta\Delta\left(\frac{\beta}{\tilde{m}_{z,i,t}}\sum_{j=1}^{\tilde{m}_{z,i,t}}\mathcal{I}\left(|z_{i,t-1} - z_{j,t-1}| \leq r_2\right)x_{j,t-1}\right)\right)\right) = 0, \quad t = 3, ..., T, \quad s = 1, .., t-2.$$

where $\tilde{m}_{z,i,t} = \min(m_{z,i,t}, m)$.

The extension presented in (22) is very important. While it is intuitive that it is likely that there exists some variable which can be used to order units (denoted by $z_{i,t}$ in (22)), it is not clear why one would want to set $z_{i,t} = x_{i,t}$ as we did in the first version of the model we presented in (1). A main reason for us doing so, in the first instance, was because then the model is self-contained and can be analysed along the lines used in section 2. But there is another reason why one may wish to focus on (1) rather than the more general (22). To see why, let us provide a simple analogy in terms of simple univariate time series models before analysing the case at hand. Let

$$x_t = s_t + u_t$$

where

$$s_t = \gamma s_{t-1} + v_t$$

and u_t and v_t are serially uncorrelated. Then, it is straightforward to see that a good approximation for this model can be provided by fitting an $AR(1)$ model to x_t. Similarly, let the true model for $x_{i,t}$ be given by a slight variation of (22) of the form

$$x_{i,t} = s_{i,t} + \epsilon_{i,t} \tag{23}$$

where

$$s_{i,t} = \frac{\beta}{m_{z,i,t}}\sum_{j=1}^{N}\mathcal{I}\left(|z_{i,t-1} - z_{j,t-1}| \leq r_2\right)q_{j,t-1} \tag{24}$$

and let

$$z_{i,t} = \gamma z_{i,t-1} + v_{i,t}$$

and

$$q_{i,t} = \delta q_{i,t-1} + \xi_{i,t}$$

By the fact that the $z_{i,t}$ and $q_{i,t}$ are serially correlated, it follows that the $s_{i,t}$ are serially correlated since units which cluster together along the z dimension at time t will be more likely to cluster together along the z dimension at time $t+1$. Therefore, the serial correlation in $q_{i,t}$ will be transmitted onto $s_{i,t}$. Furthermore, units which cluster along the z dimension will tend to have more correlated $s_{i,t}$ over i. But, of course, this means that units that cluster along the z dimension will also cluster along the x dimension, in the same order as across the z dimension, since they will have $s_{i,t}$ that are more correlated across i than units which do not cluster along the z dimension. The resulting clustering along the x dimension then implies that a term of the form $\frac{\rho}{m_{z,i,t}} \sum_{j=1}^{N} \mathcal{I}\left(|x_{i,t-1} - x_{j,t-1}| \leq r_2\right) q_{j,t-1}$ will have explanatory power for $x_{i,t}$ justifying the use of model (1). So, just as (23) can be approximated by an $AR(1)$,

$$x_{i,t} = \frac{\beta}{m_{i,t}} \sum_{j=1}^{N} \mathcal{I}\left(|z_{i,t-1} - z_{j,t-1}| \leq r\right) x_{j,t-1} + \epsilon_{i,t}, \tag{25}$$

can be approximated by(1), which has an 'AR' structure in the distance/trigger variable. The usefulness of this approximation becomes more apparent if one notes the possibility of having cross-sectional averages defined through intersections of triggering events with more than one trigger variables such as

$$x_{i,t} = \frac{\beta}{m_{z,i,t}} \sum_{j=1}^{N} \mathcal{I}\left(\cap_{s=1}^{p} \left\{\left|z_{i,t-1}^{(s)} - z_{j,t-1}^{(s)}\right| \leq r_s\right\}\right) x_{j,t-1} + \epsilon_{i,t} \tag{26}$$

where $\left(z_{i,t-1}^{(1)}, ..., z_{i,t-1}^{(p)}\right)'$ is a vector of trigger variables and $\mathcal{I}\left(\cap_{s=1}^{p} \left\{\left|z_{i,t-1}^{(s)} - z_{j,t-1}^{(s)}\right| \leq r_s\right\}\right) = 1$ if and only if $\mathcal{I}\left(\left|z_{i,t-1}^{(s)} - z_{j,t-1}^{(s)}\right| \leq r_s\right) = 1$ for all s. Further, it is also clear that even if there is structural change whereby the identity of the trigger variables changes over time, the model with the 'AR' structure in the distance/trigger variable, can still approximate the true unknown and changing model.

It is reasonable to expect that there are further sources of cross-sectional dependence in panels such as those we are considering. For example, the endogenously determined cross-sectional dependence exemplified by model (1) can be coupled with exogenous cross-sectional dependence, such as common shocks arising in the macroeconomy. Such exogenous cross-sectional dependence can be modelled by linear factor structures. Further cross-sectional dependence, of the factor variety, can be introduced by considering the following extension of (1)

$$x_{i,t} = \frac{\rho}{m_{i,t}} \sum_{j=1}^{N} \mathcal{I}\left(|x_{i,t-1} - x_{j,t-1}| \leq r\right) x_{j,t-1} + \eta_{i,t} \tag{27}$$

where

$$\eta_{i,t} = \lambda_i' f_t + \epsilon_{i,t} \tag{28}$$

and f_t is an unobserved factor. The estimation of (27) is of particular interest. If the factor is serially uncorrelated, estimation of this model along the lines suggested for estimation of (1) is possible. However, if the factor is serially correlated, it is clear that $\eta_{i,t}$ and $\tilde{x}_{i,t} = \sum_{j=1}^{N} \mathcal{I}\left(|x_{i,t-1} - x_{j,t-1}| \leq r\right) x_{j,t-1}$ are correlated. Then, we suggest estimating a parametric factor model whereby the factor is modelled as a VAR process and the following state space model

$$\bar{x}_{i,t} = x_{i,t} - \tilde{x}_{i,t} = \lambda_i' f_t + \epsilon_{i,t}$$

$$f_t = A f_{t-1} + v_t$$

is estimated by pseudo-MLE using the Kalman filter. If one entertains (16) as the chosen model then estimation may be carried out by nonlinear least squares. It is interesting to consider the behaviour of this extended model. Therefore, we reconsider the model underlying the realisations reported in Figure 2, but allow for a factor which is i.i.d. and distributed as $f_t \sim t_1$. The loadings are given by $\lambda_i \sim U(0,1)$. We are explicitly aiming to introduce extreme behaviour through the factor. We consider two values of r, given by 0.9 and 0.999. The realisations from these two different values of r are reported in Figure 5. In the first case, there is clearly a single cluster but, as expected, the factor can generate abrupt shifts of all units. We see this around observation 130 and again around observation 170. Moving on to the very persistent case, yet more interesting behaviour arises. Here it is clear that big shocks attributed to the factor can lead to the destruction or creation of new clusters. For example, a shock around observation 260 leads to consolidation of three clusters into two. Conversely, the shock at observation 325 leads to the emergence of three clusters from the existing two before the shock.

The next extension relates to our view that (1) is an attempt at modelling, rather than describing, economic behaviour without recourse to assumptions such as rationality. In this respect (1) sets out a possible way in which information from the past is analysed by agents in forming their future behaviour. Such analysis on the part of agents may be far more complex, even in schematic terms than (1), without obeying any rationality assumptions. Allowing for a variety of further nonlinearities can capture such complexity in a tractable manner. Interesting kinds of nonlinearity that can play such a role include neural network type nonlinearities combined with cross-sectional averages. An example is

$$x_{i,t} = \sum_{j=1}^{q} c_j \psi(\bar{x}_{i,t-1}, \gamma_j) + \epsilon_{i,t}$$

where

$$\bar{x}_{i,t-1} = \left(\frac{1}{m_{z_1,i,t}} \sum_{j=1}^{N} \mathcal{I}\left(|z_{1,i,t-1} - z_{1,j,t-1}| \leq r_1\right) s_{1,j,t-1}, ..., \frac{1}{m_{z_k,i,t}} \sum_{j=1}^{N} \mathcal{I}\left(|z_{k,i,t-1} - z_{k,j,t-1}| \leq r_k\right) s_{k,j,t-1} \right)',$$

$(z_{1,i,t-1},...,z_{k,i,t-1})'$ and $(s_{1,i,t-1},...,s_{k,i,t-1})'$, $i=1,...,N$, are vector of observations on some unit specific variables where, of course, both vectors may include $x_{i,t-1}$. The neural network nodes, $\psi(\bar{x}_{i,t-1},\gamma_j)$, are some continuous function of $\bar{x}_{i,t-1}$, such as, e.g., the logistic, exponential or some radial basis function. RBF functions, given by $\psi(\bar{x}_{i,t-1},\bar{x}_j,\sigma_T)$, are radially symmetrical, integrable, bounded functions, $\bar{x}_j$ are referred to as the centres of the RBFs and σ_T is a sequence of constants. Examples include the Gaussian function of the form $\exp\left(-\left(\frac{||x-t_i||}{\sigma_T}\right)^2\right)$, or the multiquadric function $\left(1+\left(\frac{||x-t_i||}{\sigma_T}\right)^2\right)^{-1}$, $\sigma_T>0$, where $||.||$ denotes Euclidean distance. Such a general specification allows for very complicated interactions within groups defined in a variety of ways. Complexity is also introduced in the way cross-sectional averages are perceived and acted upon by agents, through the use of the node functions ψ.

While the dynamic nature of the model given by (1) is interesting, it may not be able to capture contemporaneous cross-sectional dependence effects that might be very important in fields, such as financial asset pricing, where dynamics may be less prevalent, at least in terms of the conditional mean. For example, the CAPM specifies that individual asset excess returns depend contemporaneously on a market excess return index which of course can be viewed as an aggregate of individual excess returns. Alternatively, one can think of opinions (e.g., fund manager opinions) on variables such as asset return prospects, as being determined contemporaneously by agents considering the opinions of similar agents. This motivates the following extension of our basic model

$$x_{i,t}=\frac{\rho_0}{m_{0,i,t}}\sum_{j=1,j\neq i}^{N}\mathcal{I}\left(|x_{i,t}-x_{j,t}|\leq r_0\right)x_{j,t}+\frac{\rho_1}{m_{1,i,t}}\sum_{j=1}^{N}\mathcal{I}\left(|x_{i,t-1}-x_{j,t-1}|\leq r_1\right)x_{j,t-1}+\epsilon_{i,t}, \quad (29)$$

where $m_{0,i,t}$ and $m_{1,i,t}$ are defined in an obvious way. This extended model incorporates a complex mechanism for the determination of x_t since each $x_{j,t}$ depends in a complicated way on every other $x_{j,t}$. The complex nature of this extension can be best understood by noting that simulating (29) involves solving N nonlinear simultaneous equations at each point in time, where the nonlinearity has discontinuities arising from the threshold nature of the relevant functions. This is a non-trivial mathematical problem. A linear simplification may help clarify further the issue. A simplified linear version of (29) is given by

$$x_{i,t}=\frac{\rho_0}{N}\sum_{j=1}^{N}x_{j,t}+\frac{\rho_1}{N}\sum_{j=1}^{N}x_{j,t-1}+\epsilon_{i,t}$$

In the case where $\rho_1=0$, the model decouples temporally and the solution at each point in time is given by

$$x_t=\left(I-\frac{\rho}{N}\iota\iota'\right)^{-1}\epsilon_t$$

where $x_t = (x_{1,t}, ..., x_{N,t})'$, $\epsilon_t = (\epsilon_{1,t}, ..., \epsilon_{N,t})'$ and $\iota = (1, ..., 1)'$. It is worth noting that $\left(I - \frac{\rho}{N}\iota\iota'\right)^{-1}$ does not exist when $\rho = 1$.

The final extension generalises further the gamut of weighted averages that can inform the evolution of agent opinion formation or agent actions to a very general class of models which takes the form

$$x_{i,t} = \frac{\rho}{m_{i,t}^{\mathcal{S}}} \sum_{j=1}^{N} I\left(j \in \mathcal{S}_{i,t-1}\right) x_{j,t} + \epsilon_{i,t}, \tag{30}$$

where $\mathcal{S}_{i,t-1}$ denotes a set of unit indices for unit i at time $t-1$ and $m_{i,t}^{\mathcal{S}} = \sum_{j=1}^{N} I\left(j \in \mathcal{S}_{i,t-1}\right)$. This enables a wide variety of modelling options such as the existence of a leader unit or set of units whose behaviour is mimicked by other units. For example, a specific instance of (30), where

$$\mathcal{S}_{i,t-1} = \mathcal{S}_{t-1} = \arg \max_{j=1,\dots,N} \sum_{s=1}^{p} q_{j,t-s} \tag{31}$$

can be used to model fund managers that follow the best performing manager in the near past. In this case $x_{i,t}$ denotes the holdings of a given asset by manager i at time t, while $q_{i,t}$ denotes a performance measure of manager i at time t. Of course, multivariate extensions to describe the evolution of holdings for multiple assets are obvious. Similarly

$$\mathcal{S}_{i,t-1} = \mathcal{S}_{t-1} = median(\sum_{s=1}^{p} q_{j,t-s}) \tag{32}$$

can be used to proxy the behaviour of fund managers that conform to forms of benchmarking. Obviously schemes such as (31) or (32) imply a factor like covariance matrix for $x_{i,t}$. Note that specifications such as (31) or (32) are significantly different to schemes that specify a priori units that are dominant such as, e.g., macroeconometric panel models that give a leading status to US variables. The present specifications describe a *mechanism* that allocates leader status to a given unit or set of units endogenously.

This extension completes the set of extensions that we think are both interesting and relevant for the effects we attempt to capture through our basic model (1). In section 5 we will report some Monte Carlo results on the performance of the estimators suggested in this section.

4 Testing Linearity

In this section, we discuss how to test if the data support the nonlinear representation contained in the proposed models. We start by recalling what parameter values imply linearity both for the basic model (1) and the leading case of the smooth version of the model given by (16) where $w(x; \gamma) = \exp(-\gamma x^2)$.

As we noted in section 2, setting $r = 0$ reduces (1) to the panel autoregression (3), while setting $r = \infty$ gives the model (4). Both are linear models. We also see that these two linear models are nested in (16). Setting $\gamma = 0$, gives (4), whereas setting $\gamma = \infty$, gives (3). As a result and unlike standard time series models there is no unique test of linearity. Which test one carries out depends very much on which null hypothesis is of greater interest.

The differences with linearity tests for standard nonlinear time series models do not stop here. A well-known problem with linearity testing in time series relates to the fact that because there invariably exist underidentified nuisance parameters, the test statistics do not have standard distributions. For example, when two regime threshold (TAR) models are considered, the specifications usually include two autoregressive parameters and the threshold. Linearity is obtained by setting the two autoregressive parameters equal to each other, in which case the threshold parameter is not identified under the null. Further, in the case of threshold models, the problem is compounded by the fact that the threshold parameter does not have a standard asymptotic distribution in any case.

A cursory analysis of the panel threshold model suggests that no underidentified parameter problem arises here. Both linear models nested by the nonlinear models, (1) and (16), have the same number of parameters as the nonlinear models, apart from the actual parameter being restricted by the null hypothesis. As a result, testing in the context of the panel model is considerably easier. In the case of (16) and using Theorem 4, one can use the normal asymptotic approximation to carry out testing for null hypotheses relating to γ.

Testing in the context of the threshold model is more difficult due to the nonstandard distribution of $\hat{r}$. Although, we have not established this distribution, the results of Chan (1993) suggest that it should be nonstandard and very difficult to use in practice. Note that, for standard time series TAR models, the standard bootstrap has been shown to be invalid for the threshold parameter by Yu (2009), while the parametric bootstrap has been shown to be valid by Yu (2007). But, since our model is likely to suffer from a number of potential misspecifications, which would invalidate the use of the parametric bootstrap, we suggest a simulation approach for carrying out inference for this parameter, and, in particular, the use of subsampling, following Gonzalo and Wolf (2005). That paper suggests subsampling, for inference in threshold models. Subsampling has been introduced by Politis and Romano (1994) and is similar, in a number of respects, to bootstrapping. The main difference is that the resamples are of a smaller dimension than the original sample. This difference makes subsampling more robust. Subsampling is valid for the overwhelming majority of cases where the bootstrap is invalid, as discussed in Politis, Romano, and Wolf (1999).

In our case, the application of subsampling carries added complications, introduced by the fact that our sample grows in two dimensions. Following Politis, Romano, and Wolf (1999) and Kapetanios (2010), we suggest the following algorithm for creating the subsamples: Set the temporal and cross-sectional subsample sizes to $b_T = T^{\zeta}$ and $b_N = N^{\zeta}$, respectively, for some $0 < \zeta < 1$. Construct initial subsamples by sampling blocks of data temporally. These are given by $\{\tilde{x}_{1,b_T}, \tilde{x}_{2,b_T+1}, ..., \tilde{x}_{T-b_T+1,T}\}$ where $\tilde{x}_{t_1,t_2} = (x_{t_1}, ..., x_{t_2})'$. Then, for each $\tilde{x}_{t_1,t_2}$, randomly select b_N cross-sectional units to construct the B-th subsample, x_{t_1,t_2}, $t_1 = 1, .., T - b_T + 1$, $t_2 = b_T, ..., T$, $B = 1, .., T - b_T + 1$. Note that the cross-sectional units can be different across subsamples. Although this is of no importance theoretically, it makes sense to make use of information contained in as many cross-sectional units, as possible, when subsampling. ζ is a tuning parameter related to block size. No theory exists on its determination but usual values are 0.8 or 0.7. Once the subsamples have been created, r is estimated for each subsample. The empirical distribution of the set of estimates, denoted by $\hat{r}^{*,(i)}$, $i - 1, ...B$, can then be used for inference. This empirical distribution is given by

$$L_{b_T,b_N}(x) = \frac{1}{B} \sum_{s=1}^{B} 1 \left\{ b_N b_T \left(\hat{r}^{*,(s)} - \hat{r} \right) \leq x \right\}. \tag{33}$$

The following theorem justifies the use of subsampling for the nonlinear panel threshold model.

Theorem 5 *Let Assumption 1 hold for $\epsilon_{i,t}$ in (1). Then, as long as $|\rho| < 1$, $L_{b_T,b_N}(x)$ is a consistent estimate of $\Pr_P \left(NT \left(\hat{r} - r^0 \right) \leq x \right)$ where P denotes the unknown joint probability distribution of the idiosyncratic errors $\epsilon_{i,t}$..*

As a final point it is worth noting some cases where the need for testing arises for reasons that are specific to the panel nature of the model. One such leading case is when one wishes to use this model to draw inference for aggregate variables. Let $\bar{x}_t = \frac{1}{N} \sum_{j=1}^{N} x_{j,t}$. Further, consider the case where the model is of the form (3) but with the presence of an exogenous factor. This model is given by

$$x_{i,t} = \rho x_{i,t-1} + \eta_{i,t} \tag{34}$$

where $\eta_{i,t}$ is given by (28). Then, it follows that

$$\bar{x}_t = \rho \bar{x}_{t-1} + \frac{1}{N} \sum_{j=1}^{N} \eta_{i,t} = \rho \bar{x}_{t-1} + \left(\frac{1}{N} \sum_{j=1}^{N} \lambda_i' \right) f_t + \frac{1}{N} \sum_{j=1}^{N} \epsilon_{i,t} \tag{35}$$

Assuming that λ_i does not have zero mean and that $\epsilon_{i,t}$ are zero mean and i.i.d. across i, the above implies that $\bar{x}_t$ accepts a linear $AR(1)$ representation whose error tends to f_t as $N \to \infty$. Similarly, letting the model be of the form (4), but allowing for factors, gives

$$x_{i,t} = \rho \frac{1}{N} \sum_{j=1}^{N} x_{j,t-1} + \eta_{i,t} \tag{36}$$

where again $\eta_{i,t}$ is given by (28). Then,

$$\bar{x}_t = \rho \frac{1}{N} \sum_{j=1}^{N} \left(\frac{1}{N} \sum_{j=1}^{N} x_{j,t-1} \right) + \frac{1}{N} \sum_{j=1}^{N} \eta_{i,t} = \rho \bar{x}_{t-1} + \left(\frac{1}{N} \sum_{j=1}^{N} \lambda_i' \right) f_t + \frac{1}{N} \sum_{j=1}^{N} \epsilon_{i,t}$$

which, under the same assumptions as for (35), again implies that $\bar{x}_t$ accepts a linear $AR(1)$ representation whose error term tends to f_t as $N \to \infty$. This essentially justifies the widespread use of autoregressive modelling of aggregate variables. But, if the basic model for $x_{i,t}$ is given by (1), there is no justification for a linear AR model for the aggregate variable. Further, and this has more general and important implications for the modelling of the aggregate variable, if (1) holds then the aggregate variable cannot be modelled in terms of lags of the aggregate variable alone. The constituents of the aggregate variable enter the aggregate equation in complicated ways which may imply that an appropriate model of the aggregate variable is based on modelling the whole panel, even if one only cares about the aggregate variable. Therefore, a test of linearity is crucial in determining the model employed on aggregate variables.

5 Monte Carlo Study

In this section we undertake a detailed Monte Carlo study of the new model and a number of its extensions. The Monte Carlo study focuses on the small sample properties of the estimators of the nonlinear model, and does not consider the properties of the model estimator to misspecification.

5.1 Monte Carlo setup

We consider three different sets of Monte Carlo experiment. The first focuses on the main model given by (1). The second considers (9), while the third uses (16). Of course, given the number of extensions considered in the previous section, many more Monte Carlo experiments could be considered but we feel that these three give a crucial and informative snapshot of the performance of the estimators, we discussed. They enable one to have some confidence in the fact that estimation of the model can be carried out effectively with relatively small samples.

The first set of experiments uses (1), where we set $\rho = 0.9$, $r = 0.5$ and $\sigma^2_{\epsilon_i} = 0.5$. $\epsilon_{i,t} \sim N.I.I.D.(0, \sigma^2_{\epsilon_i})$. We let $N, T = 5, 10, 20, 50, 100, 200$. The grid for determining r is $0.10, 0.11, 0.12, ..., 1.09, 1.10$. The second set of experiments is like the first, but we set $\eta_i \sim N.I.I.D.(0, 1)$ and use within group estimation which simply involves demeaning both RHS and LHS variables prior to applying least squares. Finally, the third set of experiments uses the model given by (16) where $w(x, \gamma) = e^{-\gamma x^2}$ and $\gamma = 0.5$. The rest of the settings are as with the first set of experiments. The estimation method used is nonlinear least squares. We carry out 1000 replications for all experiments. The bias and variance of the estimators over the Monte Carlo replications (multiplied by 100) are reported in Tables 4-6.

5.2 Monte Carlo results

Results make very interesting reading. We start by examining the results for the first set of experiments, reported in Table 4. We look at the estimator for ρ first. The biases for this estimator are extremely small, at less than 0.01 even for $N, T = 5$. Given the very small size of the bias it is not surprising to note that there is little in terms of a clear pattern as the number of observations increase. The bias does not reduce further as N increases, for small values of T, but it does reduce as either T increase or as N increases, for moderate and large values of T. Overall, for the largest sample size ($N, T = 200$) the bias is negligible. The variance of $\hat{\rho}$ is reduced at equal rates when either N or T increases, as we expect from Theorem 2. Moving on to $\hat{r}$, we note that the biases are much larger for very small sample sizes but reduce very rapidly, again consistent with our expectations given Theorem 2. The most rapid declines occur as N, T increase from their smallest settings. Both biases and variances are reduced with either N or T increasing. Overall, it is clear that even with $N, T = 10$ one can be reasonably confident that estimation of (1) can be carried out effectively.

Next, we consider results for the second set of experiments, reported in Table 5. Here, the biases related to $\hat{\rho}$ are considerably larger. The biases are reduced as both N and T rise but they are reduced much faster with T. The variances for $\hat{\rho}$ are again much larger compared to the first set of experiments but are reduced quite quickly as the number of observations increases. Moving on to $\hat{r}$, we note that unlike $\hat{\rho}$, the estimation of r is hardly affected by the presence of individual effects. If anything, the performance of the estimator is better. This is a surprising result, but as there is little work on the small sample properties of estimators of nonlinear panel models with individual effects, our prior about the performance of this estimator was not very strong.

Finally, we consider the third and final set of experiments, whose results are reported in Table 6. The biases and variances for $\hat{\rho}$ are comparable but slightly larger than those for the first set of experiments. However, the absolute performance for this estimator is very good even for very small samples, such as $N, T = 5$. Estimation of γ in very small samples is problematic. But, as long as both N and T equal or exceed 10, estimation improves greatly. The size of the bias and variance becomes comparable to that seen for r in the first two sets of experiments.

Overall, we therefore conclude that estimation of both the autoregressive coefficient and the parameters of the nonlinear terms is quite satisfactory. More importantly, the time dimension does not have to be large, in contrast to when linear time series models are estimated. This is helpful given that many panel datasets, to which this model may be applied, have a short time dimension.

6 Empirical Illustrations

In this section, we provide two empirical applications that illustrate the potential utility of the proposed modelling approach.

6.1 Stock Returns

Perhaps surprisingly, given that our model is one that models the dynamics of the conditional mean, for our first application we consider a dataset of stock returns. We motivate this as follows. Firstly, market returns are important for individual stock returns, albeit contemporaneously, in a number of theoretical models. Our model, with its emphasis on forms of cross-sectional averages, provides a useful vehicle to model them. Second, an autoregressive specification, which is a special case of our model, is used routinely as a benchmark for modelling, and especially forecasting, stock returns. Thirdly, although a linear dynamic specification has a poor track record for modelling stock returns, a common finding in the literature is that nonlinearity has a role to play in this respect (e.g., Guidolin, Hyde, McMillan, and Ono (2009)). This is a common finding when stock return indices are analysed. Given our discussion at the end of section 4, on aggregating processes that follow our model, which implies that the aggregate has a nonlinear structure, our model can offer interesting insights. Finally, as noted in Section 3, a model of the form of (1), which uses the own lag of the dependent variable to define the dimension along which the cross-sectional averaging is carried out, can approximate models which have other variables defining distance. So, in the case of returns, the model we use approximates models that may define distance in terms of industrial sector, profitability or other characteristics. As noted earlier the approximation properties of this model are likely to be retained to a certain extent even if the identity of the variables that regulate the distance undergoes structural change over time. In this sense our model is a 'reduced form' approximation for more structural explanations for cross-sectional correlations in returns.

We consider constituent stock return data from the S&P500 at a weekly frequency. The data are from 1993W1 through 2007W52. In our dataset, only 364 companies are present throughout the period and these are the ones we analyse.

We first estimate the simple nonlinear model given by (9). We estimate $\hat{\rho} = -0.0995$, $\hat{r} = 0.08$. The t-test associated with $\hat{\rho}$ is -39.37, which is extremely significant given Theorem 2. The panel R^2 associated with the model is 0.0058, which is of course extremely low, but expected, given that we analyse stock returns. The average R^2 across cross-sectional equations is 0.0063. Next, we introduce two comparator models: a panel data AR and a model where the lagged cross-sectional average is used as an explanatory variable, i.e. the nonlinear model for $r = \infty$. For the panel data AR, $\hat{\rho} = -0.066$ with t-test given by 37.08, the panel $R^2 = 0.0052$ and the average $R^2 = 0.0053$, while for the cross-sectional average model the respective numbers are: -0.107, -28.40, 0.0033 and 0.0036. The nonlinear model has better fit, as measured by the R^2, than the comparator models. Of course, the nonlinear model has an extra parameter (the threshold) which needs to be penalised. A multivariate information criterion is not possible since the dimension of the model is so large that the determinant of the covariance matrix of the residuals, needed to construct the information criterion, is found to be numerically indistinguishable from zero. We choose to construct information criteria for each cross-sectional equation, where the penalty parameter is set to $1/N$ since the threshold parameter is shared by all cross-sectional equations. Table 7 reports the proportion of companies for which each criterion chooses the nonlinear model over the two comparator models. Again we see that the nonlinear model is preferred over its comparators.

Next, we carry out a variety of tests on the residuals of the models. In particular, for every stock return series, we obtain its residuals, from the nonlinear model and the comparator models, and test them for the following: normality (Jarque-Bera test), residual serial correlation (LM test with 1 and 4 lags), ARCH effects (LM test with 1 and 4 lags) and neglected dynamic nonlinearity (Teräsvirta, Lin, and Granger (1993) RESET type test with third order polynomial approximation and 1 lag). We report the number of rejections, at the 5% significance level, in Table 8. It seems that all residuals are non-normal, as one would expect. There is some limited evidence of further serial correlation. There is significant evidence of ARCH effects. There is considerable evidence of neglected nonlinearity. It seems that the cross-sectional model displays considerably more evidence of further serial correlation compared to the other models. The most interesting finding relates to neglected nonlinearity. The nonlinear model has about 10% fewer cases of rejection than the other models. This supports the case for the presence of the effect our model is designed to pick up.

Next, we add idiosyncratic AR components to every cross-sectional equation. This makes the specification more flexible and allows for an own-lag effect whose inclusion has a compelling rationale given the existing literature. We do not consider the panel AR model in this case for obvious reasons. In this case, $\hat{\rho} = -0.083$ with t-test given by -14.81, the panel $R^2 = 0.0098$ and the average $R^2 = 0.0103$ while for the cross-sectional average model the respective numbers are: -0.049, -11.12, 0.0095 and 0.0098. Tables 9 and 10 report the respective information criteria and test results. These again make clear that the nonlinear model is preferred. In particular, the favourable evidence from the neglected nonlinearity test is, if anything, even stronger.

As a final extension we add to the model a set of macroeconomic variables commonly used in the existing literature to model stock returns. Specifically we consider: a set of US T-bill yields (3-month, 6-month, 1-year, 2-year and 10-year), oil prices (Brent crude), effective exchange rates, industrial production, unemployment rate and CPI inflation. We add a fixed effect and consider our model augmented with these macroeconomic regressors, and the two restricted versions of the model (panel AR and cross-sectional average) which, in turn, are both augmented with the set of macroeconomic variables. Then estimation reveals $\hat{\rho} = -0.1106$ and $\hat{r} = 0.06$. The t-test associated with $\hat{\rho}$ is -47.96, which is again very significant given Theorem 2. The panel R^2 associated with the model is 0.02429, which is considerably higher than previously. The average R^2 for the nonlinear model, across cross-sectional equations, is 0.02495. Looking at the two comparator models, for the panel AR $\hat{\rho} = -0.083$ with t-test given by 45.87, the panel $R^2 = 0.02366$ and the average $R^2 = 0.02385$. These results suggests that in-sample the nonlinear model improves fit by at least 4% compared to the linear panel *AR* model. For the cross-sectional average model the respective numbers are: -0.134, -34.39, 0.0207 and 0.021. Clearly, the nonlinear model has better fit as measured by the R^2 compared to this model as well. Finally, we note that, once again, nonlinearity is less prevalent in the residuals of the nonlinear model with the nonlinearity test rejecting 138 times, while the equivalent number for the panel *AR* is 153 and, for the cross-sectional average model, 148.

6.2 Inflation Expectations

In this section we consider a widely exploited dataset that can be usefully analysed with the new nonlinear panel model. This is the Survey of Professional Forecasters (SPF) carried out from 1968 to 1990 by the American Statistical Association and the NBER and, since 1990, by the Federal Reserve Bank of Philadelphia. We should expect macroeconomic forecasts, such as those from the SPF, to be correlated among forecasters and estimation of the new nonlinear panel model is instructive in determining empirically the nature of the cross-sectional dependence. In turn, this is helpful in understanding further the nature of expectation formation. As Carroll (2003) stressed, there have been few attempts to model actual expectations data. Moreover, there have been even fewer studies of expectational data at the micro-economic level. Souleles (2004), who found considerable heterogeneity across individuals, is a notable exception. Other work, more interested in the forecasting properties of these expectational data than in testing alternative models of expectation formation, has restricted attention to modelling any dependence among the agents using factor models (see Gregory, Smith, and Yetman (2001)). Therefore it does not admit the possibility of alternative ways to model dependence, such as our nonlinear model, that may offer some insight on the nature of the dependence. Determining the nature of the dependence among a panel of forecasters also has a practical importance given that Gregory, Smith, and Yetman (2001) motivate use of the mean (across forecasters) forecast as a summary statistic, to be used for policymaking etc., when there is forecast "consensus". Forecast consensus is defined as when individual forecasts are both determined by a latent variable (a factor) subject to an idiosyncratic mean zero error, and when each forecaster places the same weight on the common component. But the (linear) mean forecast is not a valid measure of consensus under the nonlinear model (see, e.g., Manski (2010)).

In our application we focus on the one-quarter ahead CPI inflation rate forecasts from the SPF. While our model, as discussed in section 2.5, can accommodate missing data given there is so much in the SPF we conduct our analysis on a subsample of regular SPF respondents. This is common practice with the SPF and indeed any forecaster panel given that respondents come and go from the survey, for various reasons, so frequently. We focus on responses for the period 1990Q1-2010Q1, a total of 81 quarters. Over this period we have records of 18 professional forecasters, giving a total of 1458 potential observations. However, there remain significant gaps in the dataset which leave a total of 1079 actual observations. We consider the simple model given by (9), with includes fixed effects, in this case. This model generalises the model of Gregory, Smith, and Yetman (2001).

The current application provides a number of challenges. Firstly, we have to deal with the considerable number of missing observations; we assume that the pattern of missing observations is random. Secondly, we wish to allow for the joint presence of a nonlinear herding mechanism of the form we advocate, as well as the possibility of a factor structure similar to that of Gregory, Smith, and Yetman (2001). To handle missing observations we use the formulation given in (7). Noting that the inflation rate data are expressed as annualised quarter-over-quarter percentage points, the threshold is estimated to be 0.99 while the estimated autoregressive coefficient is given by 0.5303 with an associated t-statistic given by 18.44. It is worth at this point noting that the use of this model has some implications for the modelling of the aggregate forecast. As we noted in section 4 it is not appropriate to assume a linear model for the aggregate forecast and research has to consider the possibility of nonlinear conditional mean models with potential ARCH structures. In that respect, the volatility associated with the spread of forecasts around the aggregate is also an important source of information at the aggregate level.

Next, we wish to consider the possibility that the model should be augmented by an exogenous factor structure such as (27)-(28). Due to the presence of missing data, we consider a different estimation approach to that suggested when the factor extension was discussed earlier. An additional advantage of the estimation method described below is that we do not need to specify a parametric model for the unobserved factor.

Specifically, we consider an EM type algorithm, whereby we initialise estimation by obtaining some factor estimate and using it as an observed variable in a model of the form

$$x_{i,t} = \frac{\rho}{m_{i,t}} \sum_{j=1}^{N} \mathcal{I}\left(|x_{i,t-1} - x_{j,t-1}| \leq r\right) x_{j,t-1} + \lambda_i' f_t + \epsilon_{i,t} \quad (37)$$

which is estimated as if the factor were observed, and then the residuals, given by

$$\hat{\epsilon}_{i,t} = x_{i,t} - \frac{\hat{\rho}}{m_{i,t}} \sum_{j=1}^{N} \mathcal{I}\left(|x_{i,t-1} - x_{j,t-1}| \leq \hat{r}\right) x_{j,t-1}$$

are used to extract a new estimate of the factor. The whole approach is iterated to convergence. The actual factor is estimated, accommodating missing observations, by introducing a second estimation loop where for a given set of observed residuals and a given pattern of missing residuals, both the factor and the missing residuals are estimated. This is done by conditioning on a factor estimate to get estimated missing residuals using the factor and estimated loadings $\hat{\lambda}_i'$. Once these estimates are obtained one can estimate a new factor estimate. This two step estimation is again iterated to convergence.

When this estimation is carried out we find minimal changes in the parameter estimates for the nonlinear model. The threshold is estimated to be 0.99 while the estimated autoregressive coefficient is given by 0.5305 with an associated t-statistic given by 18.46. This suggest that in the presence of the nonlinear cross-sectional average a factor structure is redundant.

One alternative way to see this, that is of interest independently, is to compute a measure and test of cross-sectional dependence. We use the following statistic, which is a slight modification of the sphericity test statistic of Ledoit and Wolf (2002)

$$cd(x) = \frac{1}{N} tr\left(\left(C(x) - I\right)\left(C(x) - I\right)\right)$$

where $C(x)$ denotes the estimated correlation matrix of a given dataset, x. When the data $x_{i,t}$ are used to compute cd we get $cd(x) = 14.69$, while if the residuals from the nonlinear cross-sectional average model, without factors, are used the statistic is given by 1.76. The statistic obtained when residuals, from the nonlinear cross-sectional average model with factors, are used, is 1.75. Once again the difference is minimal suggesting that our model is capable of capturing the cross-sectional dependence of the data quite well. As a final check we also consider the statistic associated with using only a factor model without a nonlinear structure. The associated statistic is then 3.30, again illustrating the superiority of the nonlinear model.

7 Conclusions

Modelling assumptions, such as full-information rational expectations, are increasingly being questioned in economics and finance in favour of bounded forms of rationality and learning, where agents interact and form their own views by looking at other agents' views. This groupthink can explain herding, as commonly observed in financial markets, for example. While the theoretical analysis of these forms of rationality has become relatively commonplace, the development of econometric techniques and models that complement theoretical developments is less developed. This paper aims to provide an econometric panel model that incorporates useful and intuitive ideas on the structure of herding behaviour in a variety of settings.

Our model can be viewed as both a new kind of panel model, as well as a model that incorporates more structural aspects, in the sense that it can be given behavioural underpinnings. From an economic point of view, the cross-sectional average structures that appear in our panel regressions have a clear interpretation as 'shortcuts' that agents may take to form views and expectations (cf. Carroll (2003)). Such an interpretation brings our work firmly within the context of the extensive literature on bounded rationality and behavioural explanations for economic behaviour like herding.

From an econometric point of view, we make a number of contributions. Our model provides, to the best of our knowledge, the first attempt to introduce endogenous cross-sectional correlation in a panel framework where typically units, while sharing commonalities, in terms of parameters, remain stochastically uncorrelated. In doing so we link a variety of literatures such as nonlinear time series analysis, factor analysis and panel data econometrics. The model has interesting and puzzling econometric features, such as nonstandard behaviour when fixed effects are introduced and when linearity, as a restricted hypothesis, is tested. We provide a large set of extensions to the simple form of the model that allow for great modelling flexibility. These extensions possibly delineate the whole class of models that can be used to fit large N, T, panel datasets, allow endogenous cross-sectional correlations and are contained within the class of models that have finite parametric representations.

References

AKERLOF, G., AND R. SHILLER (2009): *Animal spirits: How Human Psychology Drives the Economy, and Why It Matters for Global Capitalism*. Princeton University Press.

BAI, J. (2003): "Inferential Theory for Factor Models of Large Dinensions," *Econometrica*, 71, 135–173.

BAI, J., AND S. NG (2002): "Determining the Number of Factors in Approximate Factor Models," *Econometrica*, 70, 191–221.

BANERJEE, A. V. (1992): "A Simple Model of Herd Behavior," *The Quarterly Journal of Economics*, 107(3), 797–817.

BANERJEE, S., R. KANIEL, AND I. KREMER (2009): "Price Drift as an Outcome of Differences in Higher Order Beliefs," *Review of Financial Studies*, 22, 3707–3734.

BANERJEE, S., AND I. KREMER (2010): "Disagreement and Learning: Dynamic Patterns of Trade," *Journal of Finance*, Forthcoming.

BIKHCHANDANI, S., D. HIRSHLEIFER, AND I. WELCH (1992): "A Theory of Fads, Fashion, Custom, and Cultural Change in Informational Cascades," *Journal of Political Economy*, 100(5), 992–1026.

BLONDEL, V. D., J. M. HENDRICKX, AND J. N. TSITSIKLIS (2009): "Continuous Time Average Preserving Opinion Dynamics with Opinion Dependent Communications," *Working Paper, arXiv No. 0907 4662 v1.*

CANER, M., AND B. HANSEN (2004): "Instrumental Variables Estimation of a Threshold Model," *Econometric Theory*, 20, 813–843.

CARROLL, C. D. (2003): "Macroeconomic Expectations Of Households And Professional Forecasters," *The Quarterly Journal of Economics*, 118(1), 269–298.

CHAN, K. S. (1989): "A Note on the Geometric Ergodicity of a Markov chain," *Advances in Applied Probability*, 21, 702–704.

——— (1993): "Consistency and Limiting Distribution of the Least Squares Estimator of a Threshold Autoregressive Model," *The Annals of Statistics*, 21(1), 520–533.

CHAN, K. S., AND H. TONG (1985): "On the Use of the Deterministic Lyapunov Function for the Ergodicity of Stochastic Difference Equations," *Advances in Applied Probability*, 17, 666–678.

Chevillon, C., M. Massmann, and S. Mavroeidis (2010): "Inference in Models with Adaptive Learning," *Journal of Monetary Economics*, 57, 341–351.

Chudik, A., and M. H. Pesaran (2010): "Infinite Dimensional VARs and Factor Models," *Journal of Econometrics*, p. Forthcoming.

Chudik, A., M. H. Pesaran, and E. Tosetti (2009): "Weak and Strong Cross Section Dependence and Estimation of Large Panels," *Mimeo, University of Cambridge*.

Davidson, J. (1994): *Stochastic Limit Theory*, Advanced Tests in Econometrics. Oxford University Press.

Devenow, A., and I. Welch (1996): "Rational herding in financial economics," *European Economic Review*, 40(3-5), 603–615.

Gilboa, I., O. Lieberman, and D. Schmeidler (2006): "Empirical Similarity," *Review of Economics and Statistics*, 88(3), 433–444.

Gonzalo, J., and M. Wolf (2005): "Subsampling Inference in Threshold Autoregressive Models," *Journal of Econometrics*, 127, 201–224.

Gregory, A. W., G. W. Smith, and J. Yetman (2001): "Testing for Forecast Consensus," *Journal of Business and Economic Statistics*, 19, 34–43.

Guidolin, M., S. Hyde, D. McMillan, and S. Ono (2009): "Non-Linear Predictability in Stock and Bond Returns: When and Where Is It Exploitable," *International Journal of Forecasting*, 21(2), 373–399.

Hansen, B. E. (2000): "Sample Splitting and Threshold Estimation," *Econometrica*, 68, 575–603.

Hayashi, F. (2000): *Econometrics*. Princeton University Press.

Hirshleifer, D., and S. H. Teoh (2003): "Herd Behavior and Cascading in Capital Markets: A Review and Synthesis," *Review of Financial Studies*, 9, 25–66.

Jegadeesh, N., and W. Kim (2010): "Do Analysts Herd? An Analysis of Recommendations and Market Reactions," *Review of Financial Studies*, 23, 901–937.

Kapetanios, G. (2001): "Model Selection in Threshold Models," *Journal of Time Series Analysis*, 22, 733–754.

——— (2010): "A Testing Procedure for Determining the Number of Factors in Approximate Factor Models with Large Datasets," *Journal of Business and Economic Statistics*, 28, 397–409.

KOURTELLOS, A., C. M. TAN, AND T. STENGOS (2008): "Threshold Regression with Endogenous Threshold Variables," *Working Paper Series 05-08, Rimini Centre for Economic Analysis*.

KRAUSE, U. (1997): "Soziale Dynamiken mit vielen Interakteuren. Eine Problemskizze," *Modellierung und Simulation von Dynamiken mit vielen interagierenden Akteuren*, pp. 37–51.

LEDOIT, O., AND M. WOLF (2002): "Some Hypotheses Tests for the Covariance Matrix when the Dimension is Large Compared to the Sample Size," *Annals of Statistics*, 30(4), 1081–1102.

MANSKI, C. F. (2010): "Interpreting and Combining Heterogeneous Surveys," in *Forecasts, the Oxford Handbook on Economic Forecasting*, ed. by M. Clements, and D. Hendry. Oxford University Press.

NICKELL, S. J. (1981): "Biases in Dynamic Models with Fixed Effects," *Econometrica*, 49(6), 1417–26.

PESARAN, M. H. (2006): "Estimation and Inference in Large Heterogeneous Panels with a Multifactor Error Structure," *Econometrica*, 74(4), 967–1012.

PHILLIPS, P. C. B., AND H. R. MOON (1999): "Linear Regression Limit Theory for Nonstationary Panel Data," *Econometrica*, 67(5), 1057–1112.

POLITIS, D. N., AND J. P. ROMANO (1994): "Large Sample Confidence Regions Based on Subsamples Under Minimal Assumptions," *Annals of Statistics*, 22, 2031–2050.

POLITIS, D. N., J. P. ROMANO, AND M. WOLF (1999): *Subsampling*. Springer Verlag.

SCHWARZ, H., H. R. RUTISHAUSER, AND E. STIEFEL (1973): *Numerical Analysis of Symmetric Matrices*. Prentice Hall.

SOULELES, N. S. (2004): "Expectations, Heterogeneous Forecast Errors, and Consumption: Micro Evidence from the Michigan Consumer Sentiment Surveys," *Journal of Money, Credit and Banking*, 36(1), 39–72.

STOCK, J. H., AND M. W. WATSON (2002): "Macroeconomic Forecasting Using Diffusion Indices," *Journal of Business and Economic Statistics*, 20, 147–162.

Teräsvirta, T., C. F. Lin, and C. W. J. Granger (1993): "Power of the Neural Network Linearity Test," *Journal of Time Series Analysis*, 14, 209–220.

Timmermann, A. (1994): "Can Agents Learn to Form Rational Expectations? Some Results on Convergence and Stability of Learning in the UK Stock Market," *Economic Journal*, 104, 777–798.

Tong, H. (1995): *Nonlinear time series: A dynamical system approach.* Oxford University Press.

Trueman, B. (1994): "Analyst Forecasts and Herding Behavior," *Review of Financial Studies*, 7, 97–124.

Tweedie, R. L. (1975): "Sufficient Conditions for Ergodicity and Recurrence of Markov Chains on a General State Space," *Stochastic Processes Appl.*, 3, 385–403.

Yu, P. (2007): "Likelihood-based Estimation and Inference in Threshold Regression," *Working Paper, University of Auckland.*

——— (2009): "Bootstrap in Threshold Regression," *Working Paper, University of Auckland.*

Appendix

Lemmas

In what follows, we develop some theoretical results that form the basis of our analysis. As noted earlier, we aim to analyse the general case where both N and T tend to infinity. Therefore, without loss of generality we let $N(T)$ be an unspecified function of T. For notational convenience we suppress the dependence of N on T. We have the following Lemmas.

Lemma 1 *Let $\left\{\{x_{i,t}\}_{i=1}^{N}\right\}_{t=1}^{T}$ follow (1). Then, for all $N_0 \leq N$, there exists T_0 such that for all $T > T_0$, $\left\{\{x_{i,t}\}_{i=1}^{N_0}\right\}_{t=T_0}^{T}$ is geometrically ergodic and asymptotically stationary, as long as $|\rho| < 1$. Further, if $\sup_{i\leq N_0} E\left(\epsilon_{i,t}^4\right) < \infty$, then $\sup_{i\leq N_0} E\left(x_{i,t}^4\right) < \infty$.*

Proof: We can write the part of (1) relevant for $\{x_{i,t}\}_{i=1}^{N_0}$, as

$$x_t^{(N_0)} = \Phi_t^{(N_0)} x_{t-1}^{(N_0)} + \epsilon_t \tag{38}$$

where $x_t^{(N_0)} = (x_{1,t}, ..., x_{N_0,t})'$, $\epsilon_t^{(N_0)} = (\epsilon_{1,t}, ..., \epsilon_{N_0,t})$ and $\Phi_t^{(N_0)} = [\Phi_{i,j,t}]$ where

$$\Phi_{i,j,t} = \frac{\rho}{m_{i,t}} \mathcal{I}\left(|x_{i,t-1} - x_{j,t-1}| \leq r\right).$$

Then, by Theorem A1.5 of Tong (1995), using the work of Tweedie (1975), the Lemma follows if $\sup_t \lambda_{\max}(\Phi_t^{(N_0)}) < 1$, where $\lambda_{\max}(\Phi_t^{(N_0)})$ denotes the maximum eigenvalue of $\Phi_t^{(N_0)}$ in absolute value. By Schwarz, Rutishauser, and Stiefel (1973), $\sup_t \lambda_{\max}(\Phi_t^{(N_0)})$ is bounded from above by the supremum over t of the row sum norm of $\Phi_t^{(N_0)}$. But, by the definition of $m_{i,t}$ this row sum norm is equal to ρ for all t. Therefore, the result for the first part of the Lemma follows. The second part of the Lemma, follows by the discussions in Remark B of Chan (1993), Chan and Tong (1985) and Chan (1989).

Lemma 2 *Let $\left\{\{x_{i,t}\}_{i=1}^{N}\right\}_{t=1}^{T}$ be given by*

$$x_{i,t} = q_{i,t-1} + \epsilon_{i,t}$$

such that the column sum norm of the variance covariance matrix of $\epsilon_t^{(N)}$ is $O(1)$ as $N \to \infty$. The column sum norm of the variance covariance matrix of $x_t^{(N)}$ is $O(N)$ if (i) $q_{i,t-1}$ is stationary, (ii) there is $\delta > 0$ such that for all N, there exist units $i, j = 1, ..., \delta N$ such that (a) $0 < \lim_{N\to\infty} \sup_{i=1,...,\delta N} Var\left(q_{i,t-1}\right)$ and (b) $\lim_{N\to\infty} \sup_{i=1,...,\delta N} Var\left(q_{i,t-1}\right) < \infty$ and (iii) there is $\delta > 0$ such that for all N, there exist units $i, j = 1, ..., \delta N$, such that $Cov\left(q_{i,t-1}, q_{j,t-1}\right) \neq 0$. If (ii)(a) does not hold then the column sum norm of the variance covariance matrix of $x_t^{(N)}$ is $O(1)$.

Proof: The proof is immediate once the definition of the column sum norm is taken into account.

Lemma 3 *Let* $\left\{\{x_{i,t}\}_{i=1}^{N}\right\}_{t=1}^{T}$ *follow (4). The column sum norm of the variance covariance matrix of* $x_t^{(N)}$ *is* $O(1)$.

Proof: To prove this theorem we will use the second part of Lemma 2. (4) can be written as

$$x_t = \nu + \rho\iota\bar{x}_{t-1} + \epsilon_t = \nu + \rho\Phi x_{t-1} + \epsilon_t \tag{39}$$

where $\nu = (\nu_1, ..., \nu_N)'$, $\bar{x}_{t-1} = \frac{1}{N}\sum_{j=1}^{N} x_{j,t-1}$, $\Phi = \frac{1}{N}\iota\iota'$ and $\iota = (1, ..., 1)'$. Note that Φ is idempotent. This implies that

$$x_t = \frac{(1-\rho^{t-1})}{(1-\rho)}\Phi\nu + \rho^t\Phi x_0 + \epsilon_t + \Phi\sum_{i=1}^{t-1}\rho^i\epsilon_{t-i} = \rho^t\Phi x_0 + \epsilon_t + \iota\left[\frac{1}{N}\sum_{j=1}^{N}\xi_{j,t}\right] \tag{40}$$

where

$$\xi_{j,t} = \sum_{i=1}^{t-1}\rho^i\epsilon_{j,t-i}$$

But, it is straightforward to show that

$$\lim_{N\to\infty} Var\left(\frac{1}{N}\sum_{j=1}^{N}\xi_{j,t}\right) = 0$$

this proving the Lemma.

Lemma 4 *Let* $\left\{\{x_{i,t}\}_{i=1}^{N}\right\}_{t=1}^{T}$ *follow (1). The column sum norm of the variance covariance matrix of* $x_t^{(N)}$ *is* $O(N)$.

Proof: We use Lemma 2. The model can be written as

$$x_{i,t} = q_{i,t-1} + \epsilon_{i,t}$$

where

$$q_{i,t-1} = \frac{\rho}{m_{i,t}}\sum_{j=1}^{N}\mathcal{I}\left(|x_{i,t-1} - x_{j,t-1}| \le r\right)x_{j,t-1}.$$

We need to verify the three conditions of Lemma 2. Condition (i) follows from Lemma 1. Next, we establish Condition (ii). By Lemma 1 it follows that it is sufficient to show that

$$0 < \lim_{N\to\infty} Var\left(\frac{\rho}{m_{i,t}}\sum_{j=1}^{N}\mathcal{I}\left(|x_{i,t-1} - x_{j,t-1}| \le r\right)x_{j,t-1}\right), \text{ for all } j. \tag{41}$$

By Assumption 1, we know that

$$\Pr\left(\left|\epsilon_{i,t}-\epsilon_{i,t-1}\right|>r\right)>0$$

for all j and $r<\infty$. This implies that

$$\Pr\left(\left|x_{i,t}-x_{i,t-1}\right|>r\right)>0$$

for all j and $r<\infty$. From this it follows that there exists $\epsilon>0$ such that

$$\Pr\left(\left|\frac{\rho}{m_{i,t}}\sum_{j=1}^{N}\mathcal{I}\left(\left|x_{i,t}-x_{j,t}\right|\leq r\right)x_{j,t}-\frac{\rho}{m_{i,t}}\sum_{j=1}^{N}\mathcal{I}\left(\left|x_{i,t-1}-x_{j,t-1}\right|\leq r\right)x_{j,t-1}\right|>\epsilon\right)>0 \quad (42)$$

But since by Markov's inequality

$$\Pr\left(\left|\frac{\rho}{m_{i,t}}\sum_{j=1}^{N}\mathcal{I}\left(\left|x_{i,t}-x_{j,t}\right|\leq r\right)x_{j,t}-\frac{\rho}{m_{i,t}}\sum_{j=1}^{N}\mathcal{I}\left(\left|x_{i,t-1}-x_{j,t-1}\right|\leq r\right)x_{j,t-1}\right|>\epsilon\right)<$$

$$\frac{1}{\epsilon^2}E\left|\frac{\rho}{m_{i,t}}\sum_{j=1}^{N}\mathcal{I}\left(\left|x_{i,t}-x_{j,t}\right|\leq r\right)x_{j,t}-\frac{\rho}{m_{i,t}}\sum_{j=1}^{N}\mathcal{I}\left(\left|x_{i,t-1}-x_{j,t-1}\right|\leq r\right)x_{j,t-1}\right|^2$$

(42) implies (41). The final condition to be checked is Condition (iii) of Lemma 2. We need to show that there is $\delta>0$ such that for all N, there exist units $i,j=1,...,\delta N$, such that

$$E\left[\left(\frac{\rho}{m_{i,t}}\sum_{s=1}^{N}\mathcal{I}\left(\left|x_{i,t}-x_{s,t}\right|\leq r\right)x_{s,t}\right)\left(\frac{\rho}{m_{j,t}}\sum_{s=1}^{N}\mathcal{I}\left(\left|x_{j,t}-x_{s,t}\right|\leq r\right)x_{s,t}\right)\right]\neq 0 \quad (43)$$

Let $\mathcal{M}_{j,t}$ denote the set of j such that $\mathcal{I}\left(\left|x_{i,t}-x_{j,t}\right|\leq r\right)=1$. By the geometric ergodicity of $x_t^{(N_0)}$ for all N_0, established in Lemma 1, and the fact that the stationary density of $x_t^{(N_0)}$ is strictly positive over all compact sets in $\mathbb{R}^{N_0}$ for all N_0, which is implied by our assumption that the density of $\epsilon_t^{(N_0)}$ is strictly positive over all compact sets in $\mathbb{R}^{N_0}$ for all N_0, we have that there is a non-zero proportion of units, that lie in both $\mathcal{M}_{i,t}$ and $\mathcal{M}_{j,t}$ for a non-zero proportion of j,$k=1,...,N$. This implies that (43) holds for some $\delta>0$ and units $i,j=1,...,\delta N$, proving the result of the Lemma.

Lemma 5 *Let* $\left\{\{x_{i,t}\}_{i=1}^{N}\right\}_{t=1}^{T}$ *follow (1). Then,*

$$\sup_i Var\left(\frac{\rho}{m_{i,t}}\sum_{j=1}^{N}\mathcal{I}\left(\left|x_{i,t-1}-x_{j,t-1}\right|\leq r\right)x_{j,t-1}\right)=O(1) \quad (44)$$

and

$$\inf_i Var\left(\frac{\rho}{m_{i,t}}\sum_{j=1}^{N}\mathcal{I}\left(\left|x_{i,t-1}-x_{j,t-1}\right|\leq r\right)x_{j,t-1}\right)=O(1) \quad (45)$$

Proof: We examine (44) which involves simply a form of cross-sectional averaging. (45) can be analysed similarly. It is easy to see that

$$Var\left(\frac{1}{m_{i,t}}\sum_{j\in\mathcal{M}_{i,t}}x_{j,t-1}\right)\sim\frac{1}{m_{i,t}^2}\left(\sum_{j\in\mathcal{M}_{i,t}}\sigma_{x_j}^2+2\sum_{j\in\mathcal{M}_{i,t}}\sum_{k\in\mathcal{M}_{i,t}}\sigma_{x_j,x_k}\right)$$

where $\sim$ denotes equality in order of magnitude and $\sigma_{x_j}^2$ and σ_{x_j,x_k} denote the variance of $x_{j,t-1}$ and the covariance of $x_{j,t-1}$ and $x_{k,t-1}$ respectively. The result of the Lemma follows immediately by Lemma 2.

Lemma 6 *Let $\left\{\{x_{i,t}\}_{i=1}^N\right\}_{t=1}^T$ follow (15). Then, for every $N_0 \le N$, there exists T_0 such that for all $T > T_0$, $\left\{\{x_{i,t}\}_{i=1}^{N_0}\right\}_{t=T_0}^T$ is geometrically ergodic and asymptotically stationary, as long as $|\rho| < 1$.*

Proof: As in the proof of Lemma 1, we can write part of (1) relevant for $\{x_{i,t}\}_{i=1}^{N_0}$, as

$$x_t = \tilde{\Phi}_t^{(N_0)}x_{t-1} + \epsilon_t \tag{46}$$

where $x_t^{(N_0)}$, $\epsilon_t^{(N_0)} = (\epsilon_{1,t}, ..., \epsilon_{N_0,t})$ are as in the proof of Lemma 1 and $\tilde{\Phi}_t^{(N_0)} = [\tilde{\Phi}_{i,j,t}]$ where

$$\tilde{\Phi}_{i,j,t} = \begin{cases} \frac{\rho}{\tilde{m}_{i,t}}x_{j,t-1}, \text{ if } x_{j,t-1} \in \left\{x_{1,t-1}^{(i)}, ..., x_{\tilde{m}_{i,t},t-1}^{(i)}\right\} \\ 0, \text{ otherwise} \end{cases},$$

Then, again by Theorem A1.5 of Tong (1995) the Lemma follows if $\sup_t \lambda_{\max}(\tilde{\Phi}_t^{(N_0)}) < 1$, where $\lambda_{\max}(\tilde{\Phi}_t^{(N_0)})$ denotes the maximum eigenvalue of $\tilde{\Phi}_t^{(N_0)}$ in absolute value. By Schwarz, Rutishauser, and Stiefel (1973), $\sup_t \lambda_{\max}(\tilde{\Phi}_t^{(N_0)})$ is bounded from above by the supremum over t of the row sum norm of $\tilde{\Phi}_t^{(N_0)}$. But, by the definition of $\tilde{m}_{i,t}$, this row sum norm is equal to ρ for all t. Therefore, the result follows.

Lemma 7 *Let $\left\{\{x_{i,t}\}_{i=1}^N\right\}_{t=1}^T$ follow (1). Let $m\rho < 1$. The column sum norm of the variance covariance matrix of x_t is $O(1)$.*

Proof: We denote $\tilde{\Phi}_t = \tilde{\Phi}_t^{(N)}$. We have the following MA representation of x_t.

$$x_t = \left(\prod_{j=1}^t \tilde{\Phi}_{t-j}\right)x_0 + \epsilon_t + \sum_{i=1}^{t-1}\left(\prod_{j=1}^{i-1}\tilde{\Phi}_{t-j}\right)\epsilon_{t-i} \tag{47}$$

By Lemma 6,

$$\left\|\prod_{j=1}^t \tilde{\Phi}_{t-j}\right\| = O_{a.s.}\left(\rho^t\right) \tag{48}$$

Noting that ϵ_t is an i.i.d. sequence gives,

$$E\left(x_t x_t'\right) = \Sigma_\epsilon + \sum_{i=1}^{t-1} E\left(\left(\prod_{j=1}^{i-1} \tilde{\Phi}_{t-j}\right) \epsilon_{t-i}\epsilon_{t-i}' \left(\prod_{j=1}^{i-1} \tilde{\Phi}_{t-j}\right)'\right) \tag{49}$$

Then,

$$\left\|E\left(x_t x_t'\right)\right\|_c \leq C + \sum_{i=1}^{t-1} \left\|E\left(\left(\prod_{j=1}^{i-1} \tilde{\Phi}_{t-j}\right) \epsilon_{t-i}\epsilon_{t-i}' \left(\prod_{j=1}^{i-1} \tilde{\Phi}_{t-j}\right)'\right)\right\|_c \tag{50}$$

for some constant C. We examine the first term of the sum on the RHS of (50). Denoting by $\widetilde{\mathcal{M}}_{j,t}$, the set of j such that $x_{j,t-1} \in \left\{x_{1,t-1}^{(i)}, ..., x_{\tilde{m}_{i,t},t-1}^{(i)}\right\}$, we have that

$$\tilde{\Phi}_t \epsilon_{t-1} = \left(\frac{\rho}{\tilde{m}_{i,t}} \sum_{j \in \widetilde{\mathcal{M}}_{1,t}} \epsilon_{j,t-1}, ..., \frac{\rho}{\tilde{m}_{i,t}} \sum_{j \in \widetilde{\mathcal{M}}_{N,t}} \epsilon_{j,t-1}\right)'$$

and

$$\tilde{\Phi}_t \epsilon_{t-1} \epsilon_{t-1} \tilde{\Phi}_t' =$$

$$\begin{pmatrix} \left(\frac{\rho}{\tilde{m}_{1,t}} \sum_{j \in \widetilde{\mathcal{M}}_{1,t}} \epsilon_{j,t-1}\right)^2 & \frac{\rho}{\tilde{m}_{1,t}\tilde{m}_{2,t}} \sum_{j \in \widetilde{\mathcal{M}}_{1,t}, k \in \widetilde{\mathcal{M}}_{2,t}} \epsilon_{j,t-1}\epsilon_{k,t-1} & ... & \frac{\rho}{\tilde{m}_{1,t}\tilde{m}_{N,t}} \sum_{j \in \widetilde{\mathcal{M}}_{1,t}, k \in \widetilde{\mathcal{M}}_{N,t}} \epsilon_{j,t-1}\epsilon_{k,t-1} \\ \frac{\rho}{\tilde{m}_{1,t}\tilde{m}_{2,t}} \sum_{j \in \widetilde{\mathcal{M}}_{1,t}, k \in \widetilde{\mathcal{M}}_{2,t}} \epsilon_{j,t-1}\epsilon_{k,t-1} & \left(\frac{\rho}{\tilde{m}_{2,t}} \sum_{j \in \widetilde{\mathcal{M}}_{2,t}} \epsilon_{j,t-1}\right)^2 & ... & \frac{\rho}{\tilde{m}_{2,t}\tilde{m}_{N,t}} \sum_{j \in \widetilde{\mathcal{M}}_{2,t}, k \in \widetilde{\mathcal{M}}_{N,t}} \epsilon_{j,t-1}\epsilon_{k,t-1} \\ ... & ... & ... & ... \\ \frac{\rho}{\tilde{m}_{1,t}\tilde{m}_{N,t}} \sum_{j \in \widetilde{\mathcal{M}}_{1,t}, k \in \widetilde{\mathcal{M}}_{N,t}} \epsilon_{j,t-1}\epsilon_{k,t-1} & \frac{\rho}{\tilde{m}_{2,t}\tilde{m}_{N,t}} \sum_{j \in \widetilde{\mathcal{M}}_{2,t}, k \in \widetilde{\mathcal{M}}_{N,t}} \epsilon_{j,t-1}\epsilon_{k,t-1} & ... & \left(\frac{\rho}{\tilde{m}_{N,t}} \sum_{j \in \widetilde{\mathcal{M}}_{N,t}} \epsilon_{j,t-1}\right)^2 \end{pmatrix}$$

Each set $\widetilde{\mathcal{M}}_{i,t}$ has a finite number of elements, which is bounded from above by m, uniformly over i. As a result only a finite number of the intersections $\widetilde{\mathcal{M}}_{i,t} \cap \widetilde{\mathcal{M}}_{j,t}$ are not empty which implies that only a finite number of elements of every row/column of $\tilde{\Phi}_t \epsilon_{t-1} \epsilon_{t-1} \tilde{\Phi}_t'$ have non-zero expectation. This number is bounded from above by m. As a result

$$\left\|E\left(\Phi_t \epsilon_{t-1}\epsilon_{t-1}' \Phi_t'\right)\right\|_c \leq Cm$$

By similar reasoning we can show that

$$\left\|E\left(\left(\prod_{j=1}^{i-1} \tilde{\Phi}_{t-j}\right) \epsilon_{t-i}\epsilon_{t-i}' \left(\prod_{j=1}^{i-1} \tilde{\Phi}_{t-j}\right)'\right)\right\|_c \leq C\rho^i \min(m^i, N)$$

which implies that, as long as $\frac{m}{\rho} < 1$, there exists $\kappa < 1$, such that

$$\left\|E\left(\left(\prod_{j=1}^{i-1} \tilde{\Phi}_{t-j}\right) \epsilon_{t-i}\epsilon_{t-i}' \left(\prod_{j=1}^{i-1} \tilde{\Phi}_{t-j}\right)'\right)\right\|_c \leq C\kappa^i$$

which further implies that for some constant $C > 0$,

$$\lim_{t\to\infty} \sum_{i=1}^{t-1} \left\|E\left(\left(\prod_{j=1}^{i-1} \tilde{\Phi}_{t-j}\right) \epsilon_{t-i}\epsilon_{t-i}' \left(\prod_{j=1}^{i-1} \tilde{\Phi}_{t-j}\right)'\right)\right\|_c \leq C$$

proving the lemma.

Lemma 8 *Let* $\left\{\{x_{i,t}\}_{i=1}^{N}\right\}_{t=1}^{T}$ *follow (16). Then, for every* $N_0 \leq N$*, there exists* T_0 *such that for all* $T > T_0$*,* $\left\{\{x_{i,t}\}_{i=1}^{N_0}\right\}_{t=T_0}^{T}$ *is geometrically ergodic and asymptotically stationary, as long as* $|\rho| < 1$.

Proof: Proceeding as in the proof of Lemma 1, we can write part of (16) relevant for $\{x_{i,t}\}_{i=1}^{N_0}$, as

$$x_t = \Phi_t^{w,(N_0)} x_{t-1} + \epsilon_t \tag{51}$$

where $\Phi_t^{w,(N_0)} = [\Phi_{i,j,t}^{w}]$ where

$$\Phi_{i,j,t}^{w} = \frac{\rho w(|x_{i,t-1} - x_{j,t-1}|\,;\gamma)}{\sum_{j=1}^{N} w(|x_{i,t-1} - x_{j,t-1}|\,;\gamma)}.$$

Then, the result follows along very similar lines to the proof of Lemma 1.

Lemma 9 *Let* $\left\{\{x_{i,t}\}_{i=1}^{N}\right\}_{t=1}^{T}$ *follow (16). The column sum norm of the variance covariance matrix of* $x_t^{(N)}$ *is* $O(N)$.

Proof: Let $x_t = x_t^{(N)}$and $\Phi_t^w = \Phi_t^{w,(N)}$. We have the following MA representation of x_t.

$$x_t = \left(\prod_{j=1}^{t} \Phi_{t-j}^{w}\right) x_0 + \epsilon_t + \sum_{i=1}^{t-1} \left(\prod_{j=1}^{i-1} \Phi_{t-j}^{w}\right) \epsilon_{t-i} \tag{52}$$

By Lemma 8,

$$\left\| \prod_{j=1}^{t} \Phi_{t-j}^{w} \right\| = O_{a.s.}\left(\rho^t\right) \tag{53}$$

Noting that ϵ_t is a i.i.d. sequence gives,

$$E\left(x_t x_t'\right) = \Sigma_\epsilon + \sum_{i=1}^{t-1} E\left(\left(\prod_{j=1}^{i-1} \Phi_{t-j}^{w}\right) \epsilon_{t-i}\epsilon_{t-i}' \left(\prod_{j=1}^{i-1} \Phi_{t-j}^{w}\right)'\right) \tag{54}$$

It is sufficient to show that

$$\left\|E\left(\Phi_t^w \epsilon_{t-1}\epsilon_{t-1}' \Phi_t^{w\prime}\right)\right\|_c = O(N)$$

We have that

$$\Phi_t \epsilon_{t-1} = \left(\rho \sum_{j=1}^{N} \frac{w(|x_{1,t-1} - x_{j,t-1}|\,;\gamma)}{\sum_{j=1}^{N} w(|x_{i,t-1} - x_{j,t-1}|\,;\gamma)} \epsilon_{j,t-1}, ..., \rho \sum_{j=1}^{N} \frac{w(|x_{N,t-1} - x_{j,t-1}|\,;\gamma)}{\sum_{j=1}^{N} w(|x_{i,t-1} - x_{j,t-1}|\,;\gamma)} \epsilon_{j,t-1}\right)'$$

and it follows that every element of $\Phi_t^w \epsilon_{t-1}\epsilon_{t-1}' \Phi_t^{w\prime}$ has nonzero expectation by the geometric ergodicity of $x_t^{(N_0)}$ established in Lemma 8. As a result, $\left\|E\left(\Phi_t \epsilon_{t-1}\epsilon_{t-1}' \Phi_t'\right)\right\|_c = O(N)$, thus establishing the result of the Lemma. It can again be similarly established that for any ordering of the units and any choice of N_0, the Lemma holds for $x_t^{(N_0)}$.

Lemma 10 *Let* $\left\{\{x_{i,t}\}_{i=1}^{N}\right\}_{t=1}^{T}$ *follow (16). Then,*

$$\sup_i Var\left(\sum_{j=1}^{N}\frac{\rho w(|x_{i,t-1}-x_{j,t-1}|\,;\gamma)x_{j,t-1}}{\sum_{j=1}^{N}w(|x_{i,t-1}-x_{j,t-1}|\,;\gamma)}\right)=O(1)$$

and

$$\inf_i Var\left(\sum_{j=1}^{N}\frac{\rho w(|x_{i,t-1}-x_{j,t-1}|\,;\gamma)x_{j,t-1}}{\sum_{j=1}^{N}w(|x_{i,t-1}-x_{j,t-1}|\,;\gamma)}\right)=O(1)$$

Proof: The proof follows very similarly to that of Lemma 3.

Lemma 11 *Let* $\left\{\{x_{i,t}\}_{i=1}^{N}\right\}_{t=1}^{T}$ *follow (9).Let* $\bar{\epsilon}_{i,t}=\epsilon_{i,t}-\bar{\epsilon}_i$*, where* $\bar{\epsilon}_i=\frac{1}{T}\sum_{j=1}^{T}\epsilon_{j,t}$ *Then, there exists* T_0 *such that for all* $T>T_0$*,*

$$E\left(\left[\frac{\rho}{m_{i,t}}\sum_{j=1}^{N}\mathcal{I}\left(|x_{i,t-1}-x_{j,t-1}|\leq r\right)x_{j,t-1}\right]\left(\epsilon_{i,t}-\bar{\epsilon}_i\right)\right)=O\left(\frac{1}{NT}\right).$$

Proof: We establish the result for $r=\infty$ (i.e. the linear model given by (4)). Then, the result follows by Lemma 1 and the assumption that the stationary density of $\{x_{i,t}\}_{i=1}^{N_0}$ is positive uniformly over N_0, since this implies that there exists T_0 such that for all $T>T_0$, and uniformly over i, the number of j such that $\mathcal{I}\left(|x_{i,t-1}-x_{j,t-1}|\leq r\right)=1$ for any t, is a non-zero proportion of N_0, for all N_0.

To show the result for the linear model, let, as before, $x_t=x_t^{(N)}$. As before, (4) can be written as

$$x_t=\nu+\rho\iota\bar{x}_{t-1}+\epsilon_t=\nu+\rho\Phi x_{t-1}+\epsilon_t \tag{55}$$

where $\nu=(\nu_1,...,\nu_N)'$, $\bar{x}_{t-1}=\frac{1}{N}\sum_{j=1}^{N}x_{j,t-1}$, $\Phi=\frac{1}{N}\iota\iota'$ and $\iota=(1,...,1)'$. Note that Φ is idempotent. This implies that

$$x_t=\frac{(1-\rho^{t-1})}{(1-\rho)}\Phi\nu+\rho^t\Phi x_0+\epsilon_t+\Phi\sum_{i=1}^{t-1}\rho^i\epsilon_{t-i}=\rho^t\Phi x_0+\epsilon_t+\iota\sum_{i=1}^{t-1}\rho^i\left(\frac{1}{N}\sum_{j=1}^{N}\epsilon_{j,t-i}\right) \tag{56}$$

For simplicity, we assume that $x_0=\nu=0$. We need to show that

$$E\left(\left[\frac{\rho}{N}\sum_{j=1}^{N}x_{j,t-1}\right]\left(\epsilon_{i,t}-\bar{\epsilon}_i\right)\right)=E\left(\rho\bar{x}_{t-1}\left(\epsilon_{i,t}-\bar{\epsilon}_i\right)\right)=O\left(\frac{1}{NT}\right).$$

We have that

$$\bar{x}_{t-1}=\frac{1}{N}\sum_{j=1}^{N}\epsilon_{j,t-1}+\sum_{i=1}^{t-2}\rho^i\left(\frac{1}{N}\sum_{j=1}^{N}\epsilon_{j,t-i-1}\right)$$

Then,

$$\bar{x}_{t-1}\left(\epsilon_{i,t}-\bar{\epsilon}_i\right)=\left(\frac{1}{N}\sum_{j=1}^{N}\epsilon_{j,t-1}\right)\left(\epsilon_{i,t}-\bar{\epsilon}_i\right)+\sum_{i=1}^{t-2}\rho^i\left(\frac{1}{N}\sum_{j=1}^{N}\epsilon_{j,t-i-1}\right)\left(\epsilon_{i,t}-\bar{\epsilon}_i\right) \tag{57}$$

Looking at the expectation of the first term on the RHS of (57), we have

$$E\left(\left(\frac{1}{N}\sum_{j=1}^{N}\epsilon_{j,t-1}\right)\left(\epsilon_{i,t}-\frac{1}{T}\sum_{j=1}^{T}\epsilon_{j,t}\right)\right)=\frac{1}{NT}\sigma_{\epsilon}^{2} \tag{58}$$

For the expectation of the second term on the RHS of (57), using (58), we have

$$E\left(\sum_{i=1}^{t-2}\rho^{i}\left(\frac{1}{N}\sum_{j=1}^{N}\epsilon_{j,t-i-1}\right)\left(\epsilon_{i,t}-\frac{1}{T}\sum_{j=1}^{T}\epsilon_{j,t}\right)\right)=\frac{1}{NT}\sum_{i=1}^{t-2}\rho^{i}\sigma_{\epsilon}^{2}=\frac{\left(1-\rho^{t-1}\right)\sigma_{\epsilon}^{2}}{\left(1-\rho\right)NT}$$

which proves the result.

Lemma 12 *Let* $\left\{\{x_{i,t}\}_{i=1}^{N}\right\}_{t=1}^{T}$ *follow*

$$x_{i,t}=\nu_{i}+\frac{\rho}{\tilde{m}_{i,t}}\sum_{j=1}^{\tilde{m}_{i,t}}x_{j,t-1}^{(i)}+\epsilon_{i,t} \tag{59}$$

where $x_{j,t}^{(i)}$ *and* $\tilde{m}_{i,t}$ *are defined below (57). Let* $\bar{\epsilon}_{i,t}=\epsilon_{i,t}-\bar{\epsilon}_{i}$*, where* $\bar{\epsilon}_{i}=\frac{1}{T}\sum_{j=1}^{T}\epsilon_{j,t}$ *Then,*

$$E\left(\left[\frac{\rho}{\tilde{m}_{i,t}}\sum_{j=1}^{\tilde{m}_{i,t}}x_{j,t-1}^{(i)}\right]\left(\epsilon_{i,t}-\bar{\epsilon}_{i}\right)\right)=O\left(\frac{1}{T}\right).$$

Proof: The proof proceeds using a similar line of attack as in the proof of Lemma 11. For simplicity, set $\nu=0$. Using (47) and (46) gives

$$x_{t}=\left(\prod_{j=1}^{t}\tilde{\Phi}_{t-j}\right)x_{0}+\epsilon_{t}+\sum_{i=1}^{t-1}\left(\prod_{j=1}^{i-1}\tilde{\Phi}_{t-j}\right)\epsilon_{t-i} \tag{60}$$

For simplicity, set $x_0=0$. It is easy to see that for each unit, i, its error term $\epsilon_{i,t-j}$ will enter at every lag j. This can be formalised by the following MA representation. Define $\tilde{\Phi}_{t,-i}$ to be equal to $\prod_{j=1}^{i-1}\tilde{\Phi}_{t-j}$ but with its diagonal equal to a vector of zeros. Then, (60) may be rewritten as

$$x_{t}=\epsilon_{t}+\sum_{i=1}^{t-1}\tilde{\Phi}_{t,-i}\epsilon_{t-i}+\sum_{i=1}^{t-1}\rho^{i}\left(\tilde{m}_{t}^{(i)}\odot\epsilon_{t-i}\right)$$

where $\odot$ denotes the Hadamard product, and

$$\tilde{m}_{t}^{(i)}=\left(\prod_{j=1}^{i-1}\frac{1}{\tilde{m}_{1,t-j}},\dots,\prod_{j=1}^{i-1}\frac{1}{\tilde{m}_{N,t-j}}\right)$$

It is then easy to see that for all i, $i = 1, ..., N$,

$$\left[\frac{\rho}{\tilde{m}_{i,t}}\sum_{j=1}^{\tilde{m}_{i,t}} x_{j,t-1}^{(i)}\right](\epsilon_{i,t} - \bar{\epsilon}_i) = \left(-\sum_{j=1}^{t-2}\rho^j\left(\prod_{s=1}^{j-1}\frac{1}{\tilde{m}_{i,t-s}}\right)\epsilon_{t-j}\right)\left(\frac{1}{T}\sum_{j=1}^{T}\epsilon_{j,t}\right) + R_{i,t}$$

where $R_{i,t}$ has terms with expectation of at most $O\left(\frac{1}{T}\right)$. Focusing on the first term of the RHS of (60), we have that

$$E\left(-\sum_{j=1}^{t-2}\rho^j\left(\prod_{s=1}^{j-1}\frac{1}{\tilde{m}_{i,t-s}}\right)\epsilon_{t-j}\right)\left(\frac{1}{T}\sum_{j=1}^{T}\epsilon_{j,t}\right) = -\frac{1}{T}\sum_{j=1}^{t-2}\rho^j\left(\prod_{s=1}^{j-1}\frac{1}{\tilde{m}_{i,t-s}}\right)\sigma_\epsilon^2$$

But, since by definition $1 \leq \tilde{m}_{i,t} \leq m$, it follows that

$$\sum_{j=1}^{t-2}\rho^j\left(\prod_{s=1}^{j-1}\frac{1}{\tilde{m}_{i,t-s}}\right) = O(1)$$

and therefore the result of the Lemma holds.

Lemma 13 *Let $\left\{\{x_{i,t}\}_{i=1}^{N}\right\}_{t=1}^{T}$ follow (13). Then, for all $N_0 \leq N$, there exists T_0 such that for all $T > T_0$, $\left\{\{x_{i,t}\}_{i=1}^{N_0}\right\}_{t=T_0}^{T}$ is geometrically ergodic and asymptotically stationary, as long as $p\sum_{i=1}^{p}|\rho_s| < 1$.*

Proof: As is usual for autoregressive models with more than one lag, we write the model in companion form. So, we can write the part of (13) relevant for $\{x_{i,t}\}_{i=1}^{N_0}$, as

$$x_t^{(p,N_0)} = \Phi_t^{(p,N_0)} x_{t-1}^{(p,N_0)} + \epsilon_t^{(p,N_0)} \tag{61}$$

where $x_t^{(p,N_0)} = (x_{1,t}, ..., x_{N_0,t}, ..., x_{1,t-p}, ..., x_{N_0,t-p})'$, $\epsilon_t^{(N_0)} = (\epsilon_{1,t}, ..., \epsilon_{N_0,t}, 0, ..., 0)'$,

$$\Phi_t^{(p,N_0)} = \begin{pmatrix} \tilde{\Phi}_t^{(1,N_0)} & \tilde{\Phi}_t^{(2,N_0)} & ... & \tilde{\Phi}_t^{(p,N_0)} \\ I & 0 & ... & 0 \\ ... & ... & ... & ... \\ 0 & ... & I & 0 \end{pmatrix},$$

$\tilde{\Phi}_t^{(s,N_0)} = [\tilde{\Phi}_{i,j,t}^{(s)}]$, $s = 1, ..., p$, and

$$\tilde{\Phi}_{i,j,t}^{(s)} = \frac{\rho_s}{m_{i,t,s}}\mathcal{I}\left(|x_{i,t-s} - x_{j,t-s}| \leq r\right) x_{j,t-s}.$$

Then, similarly to the proof of Lemma 1 it is sufficient that the row sum norm of $\begin{pmatrix} \tilde{\Phi}_t^{(1,N_0)} & \tilde{\Phi}_t^{(2,N_0)} & ... & \tilde{\Phi}_t^{(p,N_0)} \end{pmatrix}$ is bounded from above by one. But for this, it sufficient that $p\sum_{i=1}^{p}|\rho_s| < 1$ proving the result.

Lemma 14 *Let $\left\{\{x_{i,t}\}_{i=1}^{N}\right\}_{t=1}^{T}$ follow (14). Then, for all $N_0 \leq N$, there exists T_0 such that for all $T > T_0$, $\left\{\{x_{i,t}\}_{i=1}^{N_0}\right\}_{t=T_0}^{T}$ is geometrically ergodic and asymptotically stationary, as long as $q\sum_{i=1}^{q}|\rho_s| < 1$.*

Proof: The proof follows along very similar lines to that of Lemma 13.

Proof of Theorem 1

We prove consistency of the least squares estimator of ρ and r. We define $x_{ij,t-s} = |x_{i,t-s} - x_{j,t-s}|$ and $\mathcal{F}_{t-1} = \sigma(x_{1,t-1}, ..., x_{N,t-1}, x_{1,t-2}, ..., x_{N,t-2}, ...)$. Recall that ρ^0 and r^0 denote the true value of ρ and r, and denote the respective expectation conditional on $\mathcal{F}_{t-1}$ by $E_{\rho,r}(.|t-1)$. We proceed as in Chan (1993). Following the proof of consistency of the threshold parameter estimates by Chan (1993), we see that three conditions need to be satisfied for consistency. Firstly, we need to show that the data $x_{i,t}$ are geometrically ergodic and hence asymptotically covariance stationary (Condition C1). Secondly, we need to show that (Condition C2)

$$E\left(x_{i,t} - E_{\rho^0,r^0}(x_{i,t}|t-1)\right)^2 < E\left(x_{i,t} - E_{\rho,r}(x_{i,t}|t-1)\right)^2 \quad \forall \rho \neq \rho^0, \;\; \forall r \neq r^0, \; i = 1, .., N, \tag{62}$$

is satisfied and, thirdly, we need to show that (Condition C3)

$$\lim_{\delta \to 0} E\left(\sup_{(\rho,r)\in B((\rho^0,r^0),\delta)} |E_{\rho^0,r^0}(x_{i,t}|t-1) - E_{\rho,r}(x_{i,t}|t-1)|\right) = 0, \tag{63}$$

where $B(a,b)$ is an open ball of radius b centered around a, is satisfied. These three conditions together imply the uniform convergence of the objective function given

$$S(\rho, r) = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(x_{i,t} - \frac{\rho}{m_{i,t}}\sum_{j=1}^{N}\mathcal{I}\left(|x_{i,t-1} - x_{j,t-1}| \leq r\right)x_{j,t-1}\right)^2$$

to the limit objective function which is the key to establishing consistency. C1 is needed for obtaining a law of large numbers needed for Claim 1 of Chan (1993), and hence for convergence of the objective function. C3 is needed for uniformity of the convergence and, finally, C2 is needed to show that the limiting objective function is minimized at the true parameter values. C1 can be seen to follow from Lemma 1. We establish C2 and C3.

For C2 we have that

$$E\left(x_{i,t} - E_{\rho^0,r^0}(x_{i,t}|t-1)\right)^2 = \sigma^2_{\epsilon_i} \tag{64}$$

Letting $m^0_{i,t} = \sum_{j=1}^{N}\mathcal{I}\left(|x_{i,t-1} - x_{j,t-1}| \leq r^0\right)$, and assuming, without loss of generality, that $r \geq r^0$, we also have

$$E\left(x_{i,t} - E_{\rho,r}(x_{i,t}|t-1)\right) = \epsilon_{i,t} + \frac{\rho^0}{m^0_{i,t}}\sum_{j=1}^{N}\mathcal{I}\left(|x_{i,t-1} - x_{j,t-1}| \leq r^0\right)x_{j,t-1} - \tag{65}$$

$$\frac{\rho}{m_{i,t}}\sum_{j=1}^{N}\mathcal{I}\left(|x_{i,t-1} - x_{j,t-1}| \leq r\right)x_{j,t-1} =$$

$$\epsilon_{i,t} + \frac{(\rho^0 - \rho)}{m_{i,t}}\sum_{j=1}^{N}\mathcal{I}\left(|x_{i,t-1} - x_{j,t-1}| \leq r^0\right)x_{j,t-1} -$$

$$\frac{\rho}{m_{i,t}}\sum_{j=1}^{N}\left(\mathcal{I}\left(|x_{i,t-1}-x_{j,t-1}|\leq r\right)-\mathcal{I}\left(|x_{i,t-1}-x_{j,t-1}|\leq r^0\right)\right)x_{j,t-1}=\epsilon_{i,t}+h_{i,t-1}$$

But, under our assumption that $\epsilon_{i,t}$ is i.i.d. across i and t, $E(\epsilon_{i,t}h_{i,t-1})=0$, thus implying that

$$E\left(\epsilon_{i,t}+h_{i,t-1}\right)^2>\sigma_{\epsilon_i}^2$$

and thereby establishing C2. For C3, we have, using (65),

$$E_{\rho^0,r^0}(x_{i,t}|t-1)-E_{\rho,r}(x_{i,t}|t-1)=\frac{(\rho^0-\rho)}{m_{i,t}}\sum_{j=1}^{N}\mathcal{I}\left(|x_{i,t-1}-x_{j,t-1}|\leq r^0\right)x_{j,t-1}- \tag{66}$$

$$\frac{\rho}{m_{i,t}}\sum_{j=1}^{N}\left(\mathcal{I}\left(|x_{i,t-1}-x_{j,t-1}|\leq r\right)-\mathcal{I}\left(|x_{i,t-1}-x_{j,t-1}|\leq r^0\right)\right)x_{j,t-1}$$

We examine the first term of the RHS of (66). By Lemma 3,

$$\frac{1}{m_{i,t}}\sum_{j=1}^{N}\mathcal{I}\left(|x_{i,t-1}-x_{j,t-1}|\leq r^0\right)x_{j,t-1}=O_{m.s.}(1)$$

and so

$$\lim_{\delta\to 0}E\left(\sup_{(\rho,r)\in B((\rho^0,r^0),\delta)}\left|\frac{(\rho^0-\rho)}{m_{i,t}}\sum_{j=1}^{N}\mathcal{I}\left(|x_{i,t-1}-x_{j,t-1}|\leq r^0\right)x_{j,t-1}\right|\right)=0$$

Moving to the second term on the RHS of (66), we have, using the fact that the stationary density of $\{x_{i,t}\}_{i=1}^{N_0}$ is positive and bounded, uniformly over N_0, which follows from Assumption 1 on the density of $\{\epsilon_{i,t}\}_{i=1}^{N_0}$, that

$$\lim_{\delta\to 0}E\left(\sup_{(\rho,r)\in B((\rho^0,r^0),\delta)}\left|\frac{\rho}{m_{i,t}}\sum_{j=1}^{N}\left(\mathcal{I}\left(|x_{i,t-1}-x_{j,t-1}|\leq r\right)-\mathcal{I}\left(|x_{i,t-1}-x_{j,t-1}|\leq r^0\right)\right)x_{j,t-1}\right|\right)\leq$$

$$\lim_{\delta\to 0}\sup_{i,j}\sup_{(\rho,r)\in B((\rho^0,r^0),\delta)}Pr(|x_{i,t-1}-x_{j,t-1}|\in(r,r^0))=0$$

proving the result.

Proof of Theorem 2

We prove the rate of convergence of $\hat{r}$ to r^0. We focus on the pooled least squares estimator. Since we know that $(\hat{\rho},\hat{r})$ is consistent, we restrict the parameter space to a neighborhood of (ρ^0,r^0), given by

$$\vartheta(\Delta)=\left\{(\rho,r)\in\Omega,\ \left|\rho-\rho^0\right|<\Delta;\ \left|r-r^0\right|<\Delta,\ 0<\Delta<1\right\}$$

It is sufficient to prove that for any ε, there exists K, such that for $(\rho, r) \in \vartheta(\Delta)$, and $r > K/(NT)$,

$$\Pr\left(S(\rho, r) - S(\rho, r^0) > 0\right) > 1 - \varepsilon. \tag{67}$$

Recall that $x_{ij,t-s} = |x_{i,t-s} - x_{j,t-s}|$. Define $Q_{ij}(r) = E\left(\mathcal{I}\left(x_{ij,t} < r\right)\right)$. By Claim 1 of Proposition 1 of Chan (1993), it follows that (67) holds if for any $\varepsilon > 0$, $\eta > 0$, there exists $K > 0$ such that for all N, T

$$\inf_{1\le i,j\le N} \Pr\left(\sup_{\Delta\ge r>K/(NT)} \left|\sum_{j=1}^{N}\sum_{t=2}^{T} \frac{\mathcal{I}\left(x_{ij,t-1} < r\right)}{NTQ_{ij}(r)} - 1\right| < \eta\right) > 1 - \varepsilon, \tag{68}$$

$$\inf_{1\le i,j\le N} \Pr\left(\sup_{\Delta\ge r>K/(NT)} \left|\sum_{j=1}^{N}\sum_{t=2}^{T} \frac{\epsilon_{i,t}\mathcal{I}\left(x_{ij,t-1} < r\right)}{NTQ_{ij}(r)}\right| < \eta\right) > 1 - \varepsilon \tag{69}$$

and

$$\inf_{1\le i,j\le N} \Pr\left(\sup_{\Delta\ge r>K/(NT)} \left|\sum_{j=1}^{N}\sum_{t=2}^{T} \frac{x_{j,t-1}\epsilon_{i,t}\mathcal{I}\left(x_{ij,t-1} < r\right)}{NTQ_{ij}(r)}\right| < \eta\right) > 1 - \varepsilon \tag{70}$$

By Claim 2 of Proposition 1 of Chan (1993), (68)- (70) hold if there exists $H < \infty$, such that

$$\sup_{1\le i,j\le N} Var\left(\sum_{j=1}^{N}\sum_{t=2}^{T} \mathcal{I}\left(x_{ij,t-1} < r\right)\right) \le NTH \sup_{1\le i,j\le N} Q_{ij}(r), \tag{71}$$

$$\sup_{1\le i,j\le N} Var\left(\sum_{j=1}^{N}\sum_{t=2}^{T} |x_{j,t-1}\epsilon_{i,t}|\,\mathcal{I}\left(r_1 < x_{ij,t-1} < r_2\right)\right) \le NTH \sup_{1\le i,j\le N} \left(Q_{ij}(r_2) - Q_{ij}(r_1)\right) \tag{72}$$

and

$$\sup_{1\le i,j\le N} Var\left(\sum_{j=1}^{N}\sum_{t=2}^{T} x_{j,t-1}\epsilon_{i,t}\mathcal{I}\left(x_{ij,t-1} < r\right)\right) \le NTH \sup_{1\le i,j\le N} Q_{ij}(r) \tag{73}$$

But, by Lemma 1and the boundedness of the indicator function, it follows that there exists $0 < m < M < \infty$ such that

$$mr \le \sup_{1\le i,j\le N} Q_{ij}(r) \le Mr \tag{74}$$

Then, by (74), the uniform boundedness of the indicator function and the second part of Lemma 1, (71)-(73) follow, thus proving the result for the rate of convergence. The second part of the theorem follows similarly to the proof of Theorem 2 and (4.11) of Chan (1993).

Proof of Theorem 3

We wish to prove that the estimator of (ρ^0, r^0, m^0), denoted by $(\hat{\rho}, \hat{r}, \hat{m})$ is consistent. Let $E_{\rho,r,m}(x_{i,t}|t-1)$ denote the expectation of $x_{i,t}$ conditional on $\mathcal{F}_{t-1}$, for a given set of parameters (ρ, r, m). Since m only takes discrete values, it is sufficient to show that

$$E\left(x_{i,t} - E_{\rho^0,r^0,m^0}(x_{i,t}|t-1)\right)^2 < E\left(x_{i,t} - E_{\rho,r,m}(x_{i,t}|t-1)\right)^2, \ \forall \rho \ne \rho^0, r \ne r^0, m \ne m^0,\ i = 1, .., N, \tag{75}$$

Then, the result follows by the proof of Theorem 1. To show (75), we have the following:

$$E\left(x_{i,t}-E_{\rho,r,m}(x_{i,t}|t-1)\right)=\epsilon_{i,t}+\frac{\rho^0}{\tilde{m}^0_{i,t}}\sum_{j=1}^{\tilde{m}^0_{i,t}}x^{(i)}_{j,t-1}-\frac{\rho}{\tilde{m}_{i,t}}\sum_{j=1}^{\tilde{m}_{i,t}}x^{(i)}_{j,t-1} \tag{76}$$

where $\tilde{m}^0_{i,t}=\min(m^0_{i,t},m^0)\quad m^0_{i,t}=\sum_{j=1}^{N}\mathcal{I}\left(|x_{i,t-1}-x_{j,t-1}|\leq r^0\right)$. As in (65), and assuming, without loss of generality, that $r\geq r^0$, (76) can be written as

$$\epsilon_{i,t}+\frac{(\rho^0-\rho)}{m^0_{i,t}}\sum_{j=1}^{\tilde{m}^0_{i,t}}x^{(i)}_{j,t-1}-\frac{\rho}{m_{i,t}}\left(\sum_{j=1}^{\tilde{m}_{i,t}}x^{(i)}_{j,t-1}-\sum_{j=1}^{\tilde{m}^0_{i,t}}x^{(i)}_{j,t-1}\right)=\epsilon_{i,t}+h_{i,t-1}$$

Then, the proof proceeds as that of Theorem 1.

Proof of Theorem 4

We wish to prove that the NLS estimator of (ρ^0,γ^0), denoted by $(\hat{\rho},\hat{\gamma})$ is consistent and asymptotically normal. For consistency, we need to establish conditions (62) and (63) but for the model given by (16). These follow along very similar lines to those for the threshold model and are therefore omitted. These conditions together with geometric ergodicity imply consistency.

For asymptotic normality, we note that using, e.g., Proposition 7.8 of Hayashi (2000), and noting that, under our assumptions, (ρ^0,γ^0) lies in the interior of the parameter space and $w(.;.)$ is twice differentiable and integrable, it is sufficient to show that

$$\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=2}^{T}\left(\sum_{j=1}^{N}\frac{\frac{\partial w}{\partial\gamma}(|x_{i,t-1}-x_{j,t-1}|\,;\gamma^0)x_{j,t-1}\epsilon_{i,t}}{\sum_{j=1}^{N}w(|x_{i,t-1}-x_{j,t-1}|\,;\gamma^0)}\right)\xrightarrow{d}N\left(0,\sigma^2_{w_i}\right),\text{ uniformly over } i. \tag{77}$$

and

$$\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=2}^{T}\left(\sum_{j=1}^{N}\frac{\rho^0w(|x_{i,t-1}-x_{j,t-1}|\,;\gamma^0)x_{j,t-1}\epsilon_{i,t}}{\sum_{j=1}^{N}w(|x_{i,t-1}-x_{j,t-1}|\,;\gamma^0)}\right)\xrightarrow{d}N\left(0,\sigma^2_{\partial w_i}\right),\text{ uniformly over } i. \tag{78}$$

We prove (77). (78) follows similarly. One way to prove the result is to show first sequential convergence in distribution, with respect to N and T and then that sequential convergence with respect to N and T, uniformly over i, implies joint convergence in distribution, uniformly over i, with respect to N and T.

We prove sequential convergence first. We examine $w_{i,j,t}\epsilon_{i,t}$ where

$$w_{i,j,t}=\sum_{j=1}^{N}\frac{\frac{\partial w}{\partial\gamma}(|x_{i,t-1}-x_{j,t-1}|\,;\gamma^0)x_{j,t-1}}{\sum_{j=1}^{N}w(|x_{i,t-1}-x_{j,t-1}|\,;\gamma^0)}.$$

By Lemma 10 which implies that $w_{i,j,t}$ has finite variance, uniformly over i, the fact that $w_{i,j,t}$ and $\epsilon_{i,t}$ are independent, and the fact that $\epsilon_{i,t}$ has finite variance, uniformly over i, by assumption, it follows that $w_{i,j,t}\epsilon_{i,t}$ is a martingale difference with finite second moments. Hence, a martingale difference CLT holds for $w_{i,j,t}\epsilon_{i,t}$ proving sequential convergence, uniformly over i.

Next we prove that sequential convergence implies joint convergence. Define

$$Y_{N,T} = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \frac{1}{\sqrt{T}} \sum_{t=2}^{T} \left(\sum_{j=1}^{N} \frac{\frac{\partial w}{\partial \gamma}(|x_{i,t-1} - x_{j,t-1}|\,; \gamma^0) x_{j,t-1}\epsilon_{i,t}}{\sum_{j=1}^{N} w(|x_{i,t-1} - x_{j,t-1}|\,; \gamma^0)} \right)$$

and

$$Y_N = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \left[\operatorname{p}\lim_{T\to\infty} \left(\frac{1}{\sqrt{T}} \sum_{t=2}^{T} \left(\sum_{j=1}^{N} \frac{\frac{\partial w}{\partial \gamma}(|x_{i,t-1} - x_{j,t-1}|\,; \gamma^0) x_{j,t-1}\epsilon_{i,t}}{\sum_{j=1}^{N} w(|x_{i,t-1} - x_{j,t-1}|\,; \gamma^0)} \right) \right) \right]$$

Sequential convergence implies that there exists Y_N such that $Y_{N,T} \xrightarrow{d} Y_N$ as $T \to \infty$.

Then, from Lemma 6 of Phillips and Moon (1999) the result follows if we show that

$$\limsup_{N,T} |E(f(Y_{N,T})) - E(f(Y_N))| = 0, \ \forall f \in \mathcal{C} \tag{79}$$

where $\mathcal{C}$ is the space of all bounded continuous real functions on $\mathbb{R}$. Without loss of generality let the functions f be such that $\left|f^{(k)}(x)\right| \leq 1$ where $f^{(k)}(x)$ denotes the k-th derivative function of $f(x)$. Fix f. Let

$$g(h) = \sup_x |f(x+h) - f(x) - f'(x)h|$$

Set $x = Y_{N,T}$ and $h = Y_{N,T} - Y_N$. It follows by the triangle inequality that

$$\limsup_{N,T} |E(f(Y_{N,T})) - E(f(Y_N))| \leq \tag{80}$$

$$\limsup_{N,T} |E(f'(Y_{N,T})(Y_{N,T} - Y_N))| + \limsup_{N,T} |E(g(Y_{N,T} - Y_N))| \tag{81}$$

But since $\left|f^{(k)}(x)\right| \leq 1$

$$\limsup_{N,T} |E(f'(Y_{N,T})(Y_{N,T} - Y_N))| \leq \limsup_{N,T} |E(Y_{N,T}) - E(Y_N)| \tag{82}$$

Also by the mean value theorem and for some finite M

$$g(h) \leq M \min\{|h|\,, h^2\}$$

Thus,

$$\limsup_{N,T} |E(g(Y_{N,T} - Y_N))| \leq M \limsup_{N,T} E|Y_{N,T} - Y_N| \tag{83}$$

From (82) and (83), it follows that the result is true if

$$\limsup_{N,T} E|Y_{N,T} - Y_N| = 0 \tag{84}$$

However, uniform integrability of $|Y_{N,T}|$, implies (84). By Theorem 12.10 of Davidson (1994) $\sup_{N,T} E|Y_{N,T}|^{\vartheta} < \infty$, for some $\vartheta > 1$ implies uniform integrability of $|Y_{N,T}|$. Hence, the result follows, by Lemma 10 and the fact that $\epsilon_{i,t}$ are assumed to have finite variance uniformly over i.

Proof of Theorem 5

Define

$$J_{T,N}(x,P) = \Pr_P \left\{ NT \left(\hat{r} - r^0 \right) \leq x \right\}. \tag{85}$$

Denote by $J(x,P)$ the limit of $J_{T,N}(x,P)$ as $N,T \to \infty$.The subsampling approximation to $J(x,P)$ is given by $L_{b_T,b_N}(x)$. For x_α, where $J(x_\alpha,P) = \alpha$, we need to prove that

$$L_{b_T,b_N}(x_\alpha) \to J(x_\alpha,P)$$

for the theorem to hold. But,

$$E(L_{b_T,b_N}(x_\alpha)) = J_{T,N}(x,P)$$

because as discussed in Section 4, the subsample is a sample from the true model, retaining the temporal ordering of the original sample. Hence, it suffices to show that $Var(L_{b_T,b_N}(x_\alpha)) \to 0$ as $N,T \to \infty$. Let

$$1_{b_T,b_N,s} = 1 \left\{ b_N b_T \left(\hat{r}^{*,(s)} - \hat{r} \right) \leq x_\alpha \right\}, \tag{86}$$

$$v_{B,h} = \frac{1}{B} \sum_{s=1}^{B} Cov \left(1_{b_T,b_N,s}, 1_{b_T,b_N,s+h} \right). \tag{87}$$

Then,

$$Var \left(L_{b_T,b_N}(x_\alpha) \right) = \frac{1}{B} \left(v_{B,0} + 2 \sum_{h=1}^{B} v_{B,h} \right) = \tag{88}$$

$$\frac{1}{B} \left(v_{B,0} + 2 \sum_{h=1}^{Cb_T - 1} v_{B,h} \right) + \frac{2}{B} \sum_{h=Cb_T}^{B} v_{B,h} = V_1 + V_2.$$

for some $C > 1$. We first determine the order of magnitude of V_1. By the boundedness of $1_{b_T,b_N,s}$, it follows that $v_{B,h}$ is uniformly bounded across h. Hence, $|V_1| \leq \frac{Cb_T}{B} \max_h |v_{B,h}|$, from which it follows that $V_1 = O(Cb_T/B) = o(1)$. We next examine V_2. For this we have that

$$|V_2| \leq \frac{2}{B} \sum_{h=Cb_T}^{B-1} |v_{B,h}|, \tag{89}$$

But, by Lemma 1, it follows that

$$v_{B,h} = o(1), \text{ uniformly across } h. \tag{90}$$

Note that this follows by the geometric ergodicity and, hence β-mixing of the process. Further, note that (90) follows for any random selection of cross sectional units undertaken to construct the subsamples. Hence,

$$\frac{2}{B} \sum_{h=Cb_T}^{B-1} |v_{B,h}| = o(1),$$

proving the convergence of $L_{b_T,b_N}(x_\alpha)$ to $J(x_\alpha,P)$.

Table 1: Proportion of variance explained by successive principal components for the nonlinear and linear models ($T = 200$, $\rho = 0.999$, $r = 0.5$(nonlinear) or $r = \infty$ (linear))

No. of PC	Nonlinear	Linear
1	0.390	0.346
2	0.520	0.359
3	0.607	0.372
4	0.662	0.385
5	0.700	0.397
6	0.728	0.409
7	0.750	0.421
8	0.767	0.433

Table 2: Proportion of variance explained by successive principal components for the nonlinear model as N rises ($T = 200$, $\rho = 0.9$, $r = 0.1$)

No. of PC/N	100	200	400	800	1000	1500
1	0.102	0.090	0.087	0.079	0.086	0.091
2	0.203	0.167	0.172	0.163	0.168	0.166
3	0.289	0.229	0.249	0.234	0.244	0.237
4	0.319	0.301	0.292	0.308	0.292	0.297
5	0.382	0.363	0.353	0.362	0.360	0.357
6	0.442	0.419	0.411	0.417	0.414	0.415
7	0.495	0.462	0.450	0.468	0.448	0.451
8	0.542	0.500	0.496	0.500	0.489	0.494

Table 3: Proportion of variance explained by successive principal components for the linear model as N rises ($T = 200$, $\rho = 0.9$, $r = \infty$)

No. of PC/N	100	200	400	800	1000	1500
1	0.037	0.036	0.022	0.017	0.016	0.015
2	0.071	0.062	0.044	0.034	0.033	0.030
3	0.105	0.088	0.065	0.050	0.049	0.045
4	0.138	0.112	0.085	0.067	0.066	0.059
5	0.170	0.136	0.104	0.083	0.082	0.074
6	0.199	0.160	0.123	0.099	0.097	0.088
7	0.228	0.183	0.142	0.114	0.113	0.102
8	0.256	0.206	0.160	0.130	0.127	0.116

Table 4: Estimation Results for threshold model. Bias and Variance

100* Sample Bias for ρ						
N/T	5	10	20	50	100	200
5	0.377	-0.138	-0.323	-0.242	-0.299	-0.181
10	0.386	0.021	-0.153	-0.165	-0.080	-0.075
20	0.749	0.218	0.033	-0.008	-0.024	-0.045
50	0.695	0.342	0.128	0.005	0.014	-0.006
100	0.760	0.294	0.130	0.039	0.021	-0.007
200	0.754	0.316	0.160	0.054	0.018	0.005
100* Sample Variance for ρ						
5	0.303	0.156	0.088	0.047	0.030	0.017
10	0.141	0.062	0.037	0.022	0.014	0.008
20	0.065	0.026	0.018	0.012	0.007	0.004
50	0.025	0.010	0.007	0.004	0.003	0.002
100	0.014	0.005	0.004	0.002	0.002	0.001
200	0.008	0.003	0.002	0.001	0.001	0.000
100* Sample Bias for r						
5	-4.740	2.401	2.980	-0.411	-1.069	-0.742
10	4.523	2.190	0.655	-1.098	-0.636	-0.195
20	3.428	-0.891	-0.976	-0.391	-0.019	-0.018
50	1.578	-1.186	-0.545	0.012	0.037	0.008
100	0.071	-0.749	-0.122	-0.079	0.003	-0.022
200	-0.277	-1.080	-0.270	-0.044	-0.032	0.004
100* Sample Variance for r						
5	10.025	7.875	5.438	2.197	1.106	0.360
10	7.028	4.659	2.604	0.729	0.192	0.052
20	5.211	2.482	1.254	0.209	0.048	0.010
50	3.026	1.265	0.431	0.061	0.021	0.003
100	2.238	0.762	0.190	0.034	0.011	0.002
200	2.102	0.539	0.149	0.028	0.007	0.001

Table 5: Estimation Results for threshold model with individual effects. Bias and Variance

100* Sample Bias for ρ						
N/T	5	10	20	50	100	200
5	-14.125	-5.127	-2.498	-1.397	-0.913	-0.588
10	-11.438	-4.405	-1.943	-1.067	-0.768	-0.467
20	-10.708	-3.926	-1.683	-0.913	-0.663	-0.478
50	-10.529	-3.637	-1.604	-0.860	-0.622	-0.456
100	-10.655	-3.748	-1.633	-0.827	-0.622	-0.438
200	-10.349	-3.666	-1.593	-0.832	-0.608	-0.451
100* Sample Variance for ρ						
5	2.476	0.465	0.123	0.043	0.019	0.011
10	1.038	0.168	0.053	0.015	0.009	0.005
20	0.446	0.072	0.019	0.007	0.004	0.003
50	0.153	0.028	0.006	0.002	0.002	0.001
100	0.092	0.015	0.004	0.001	0.001	0.001
200	0.038	0.007	0.002	0.001	0.000	0.000
100* Sample Bias for r						
5	-3.482	3.176	4.241	0.578	0.491	0.139
10	7.406	5.767	2.221	-0.535	-1.429	-0.593
20	6.797	1.357	-1.216	-0.656	-0.127	-0.067
50	3.994	0.136	-0.163	-0.027	-0.001	0.005
100	2.024	-0.288	0.040	-0.055	-0.007	0.002
200	2.239	-0.176	0.009	-0.027	0.000	0.000
100* Sample Variance for r						
5	10.304	9.240	7.948	5.542	3.761	2.714
10	8.207	5.458	3.816	2.105	1.096	0.347
20	5.469	2.801	1.436	0.371	0.120	0.027
50	3.476	1.359	0.268	0.036	0.006	0.001
100	2.785	0.547	0.080	0.007	0.001	0.000
200	2.796	0.340	0.032	0.002	0.000	0.000

Table 6: Estimation Results for smooth (exponential) model. Bias and Variance

100* Sample Bias for ρ						
N/T	5	10	20	50	100	200
5	-0.035	-0.760	-1.004	-1.115	-0.708	-0.387
10	0.097	-0.443	-0.578	-0.510	-0.339	-0.316
20	0.350	-0.296	-0.287	-0.118	-0.259	-0.153
50	-0.013	-0.275	-0.107	-0.151	-0.150	-0.103
100	-0.023	-0.290	-0.161	-0.119	-0.034	-0.109
200	-0.449	-0.435	-0.135	-0.042	-0.042	-0.062
100* Sample Variance for ρ						
5	0.513	0.303	0.249	0.124	0.093	0.058
10	0.320	0.155	0.101	0.086	0.057	0.043
20	0.142	0.073	0.062	0.046	0.039	0.031
50	0.109	0.035	0.031	0.025	0.023	0.018
100	0.101	0.024	0.018	0.015	0.012	0.012
200	0.052	0.013	0.010	0.008	0.008	0.006
100* Sample Bias for γ						
5	107.369	119.040	1.246	0.471	0.250	-0.097
10	25.474	1.751	0.804	0.376	0.120	0.116
20	4.844	1.957	0.761	0.275	0.137	0.107
50	4.925	1.474	0.675	0.247	0.240	0.112
100	5.062	1.601	0.610	0.210	0.047	0.059
200	5.271	1.904	0.484	0.158	0.069	0.030
100* Sample Variance for γ						
5	11955.80	132836.9	1.144	0.511	0.254	0.143
10	9762.673	1.900	0.420	0.256	0.153	0.085
20	3.893	0.406	0.222	0.129	0.087	0.052
50	1.920	0.168	0.099	0.063	0.046	0.027
100	1.510	0.103	0.062	0.039	0.021	0.014
200	0.463	0.063	0.032	0.019	0.013	0.007

Table 7: Stock return empirical application. Proportion of series for which information criteria choose the simple nonlinear model over comparator models over 364 series

Criterion	Panel AR	Cross-sectional Average
AIC	0.604	0.634
BIC	0.604	0.634
Hannan-Quinn	0.604	0.634

Table 8: Stock return empirical application. Test results for simple models. Proportion of series for which tests reject their respective null hypotheses over 364 series.

Test	Nonlinear model	Panel AR	Cross-sectional Average
Normality	1	1	364
LM(SC(1))	0.253	0.247	134
LM(SC(4))	0.016	0.016	0.016
LM(ARCH(1))	0.849	0.838	0.835
LM(ARCH(4))	0.489	0.491	0.491
Nonlinearity	0.393	0.426	0.428

Table 9: Stock return empirical application. Test results for models, augmented with idiosyncratic AR component. Proportion of series for which tests reject their respective null hypotheses over 364 series.

Test	Nonlinear model	Cross-sectional Average
Normality	1	1
LM(SC(1))	0	0
LM(SC(4))	0.016	0.016
LM(ARCH(1))	0.827	298
LM(ARCH(4))	0.475	0.483
Nonlinearity	0.395	0.442

Table 10: Stock return empirical application. Proportion of series for which information criteria choose the nonlinear model, augmented with idiosyncratic AR component, over comparator model over 364 series

Criterion	Cross-sectional Average
AIC	0.536
BIC	0.536
Hannan-Quinn	0.536

Figure 1: Results for the deterministic model. First panel: $r = 0.5$, Second panel: $r = 3$.

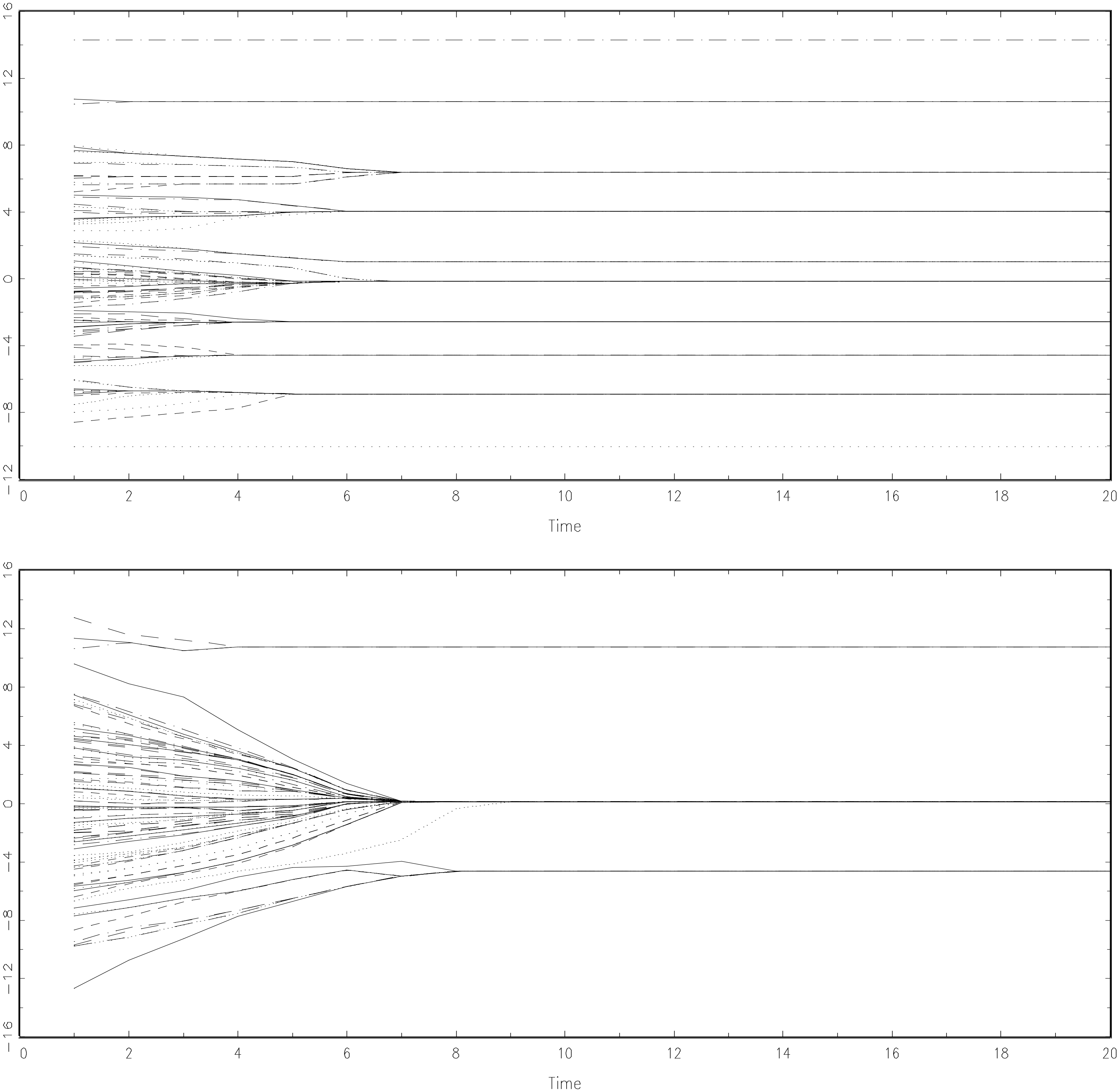
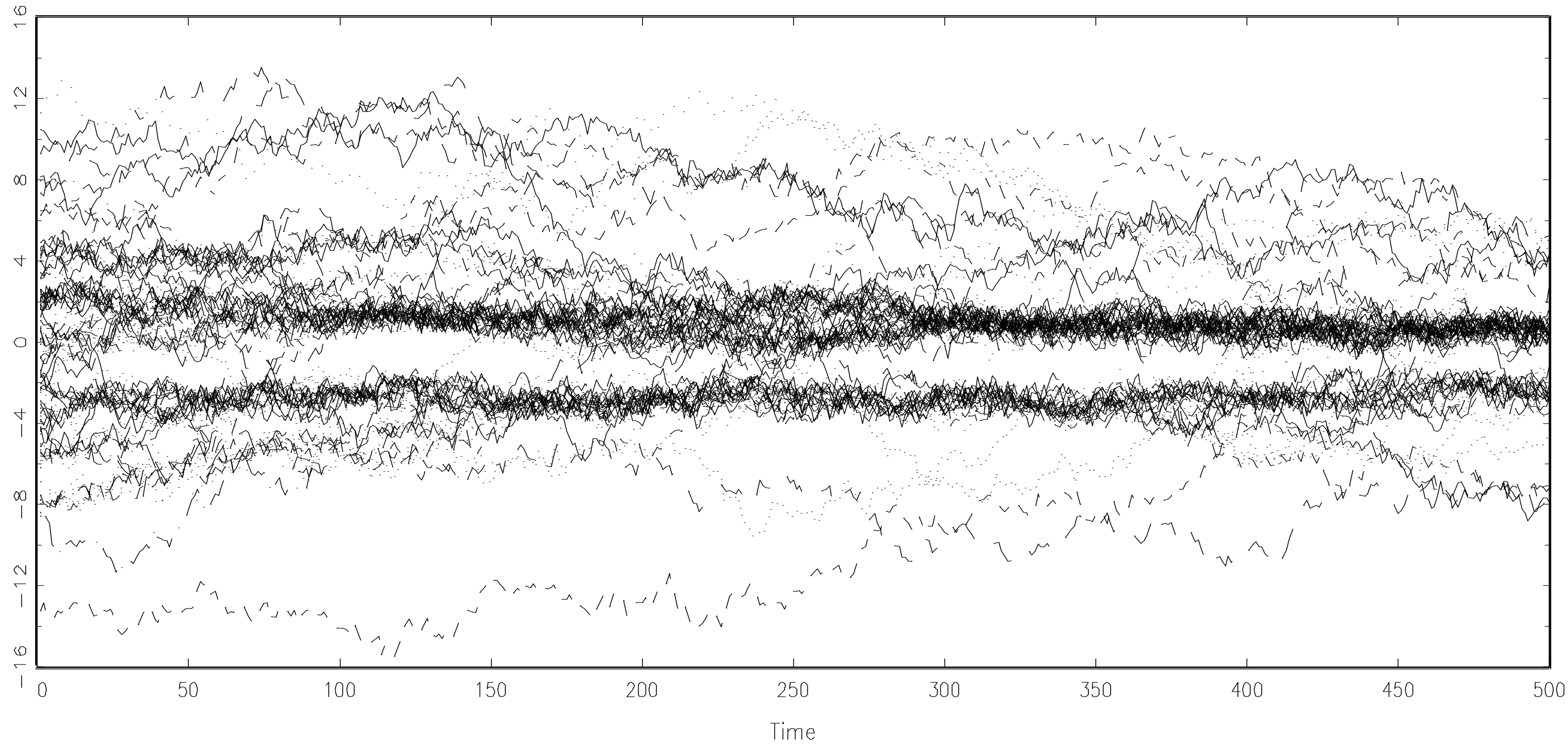


Figure 2: Results for the stochastic model. Two different realisations



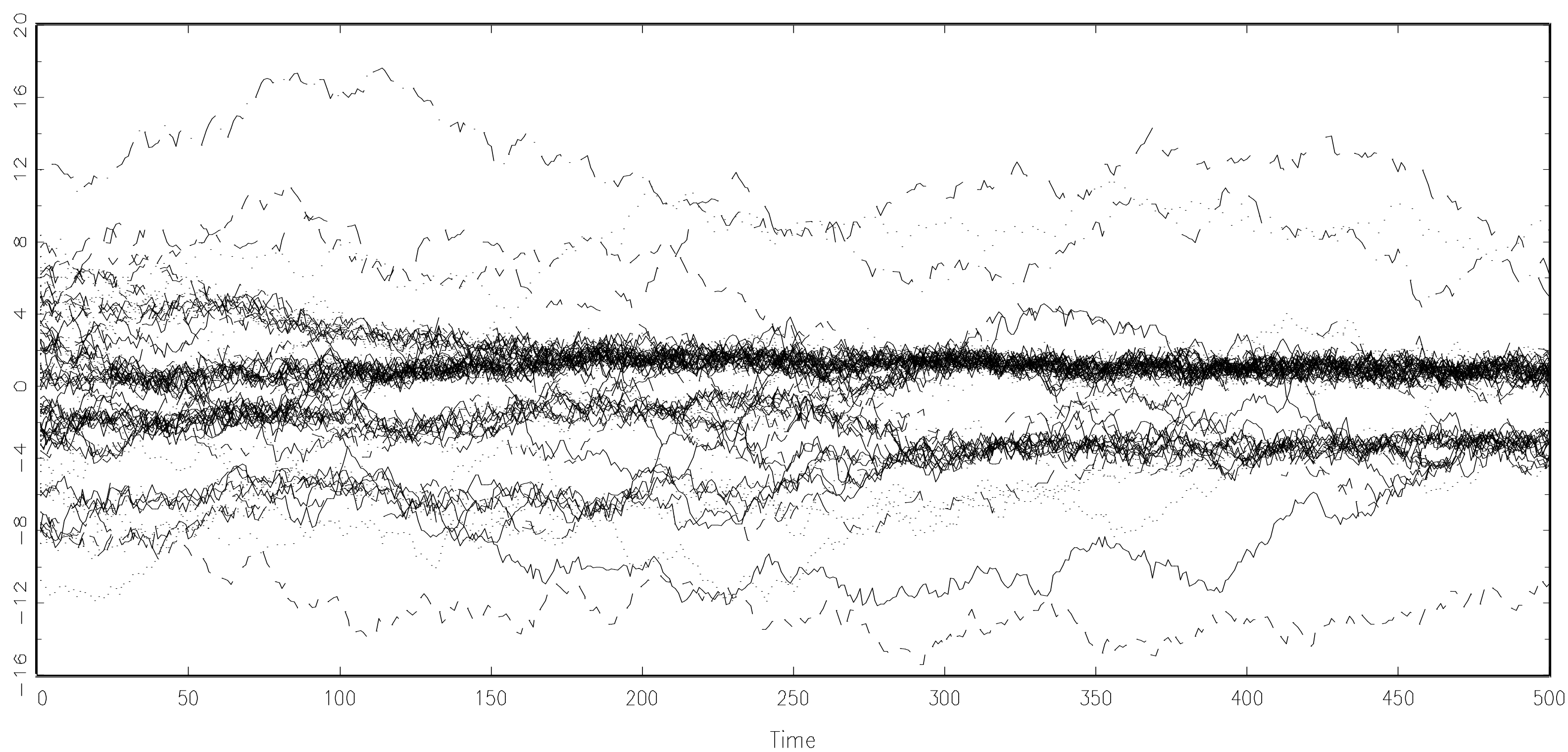


Figure 3: Results for the stochastic model using errors with fat tails

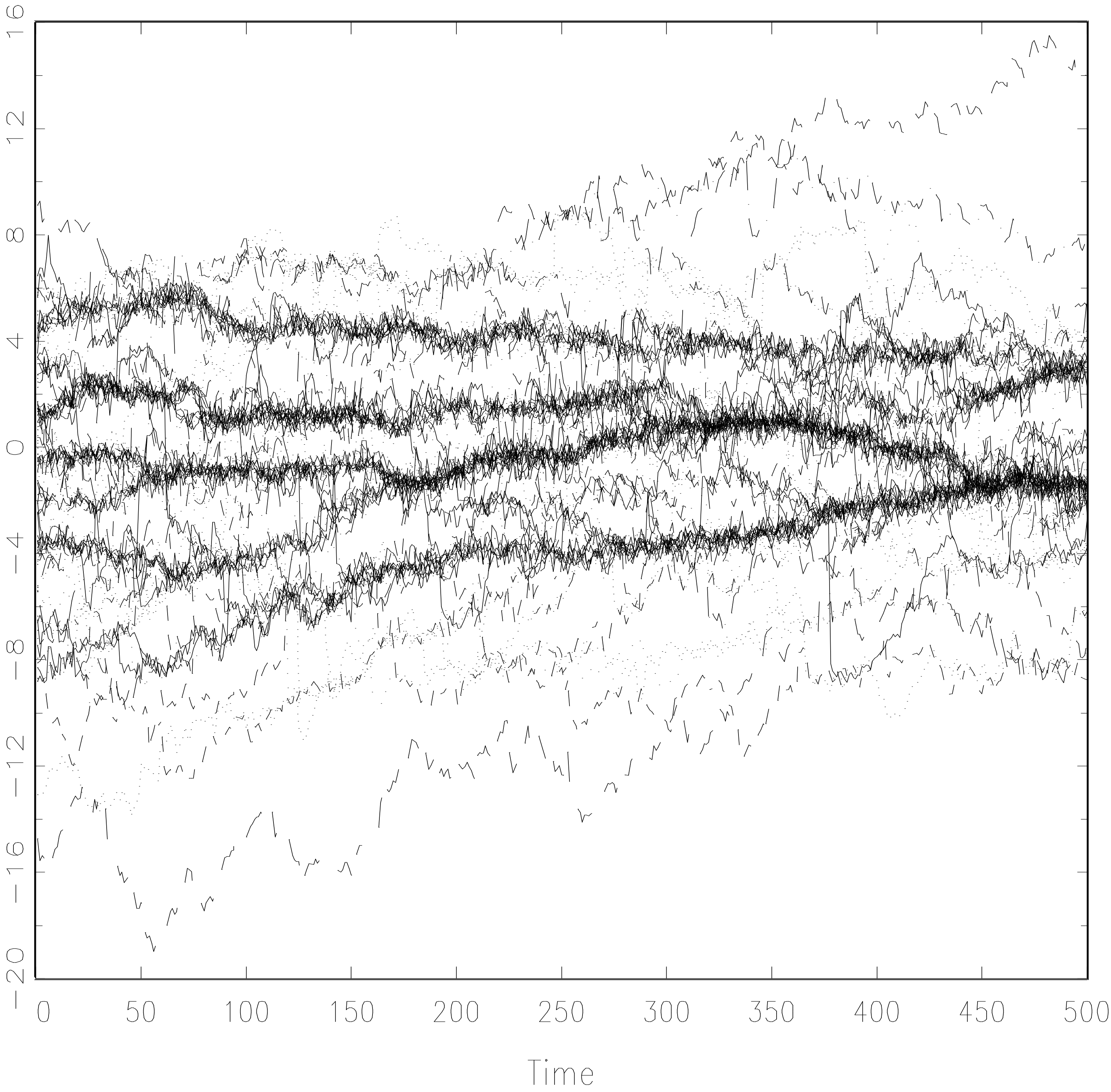


Figure 4: Results for the restricted linear version of the stochastic model

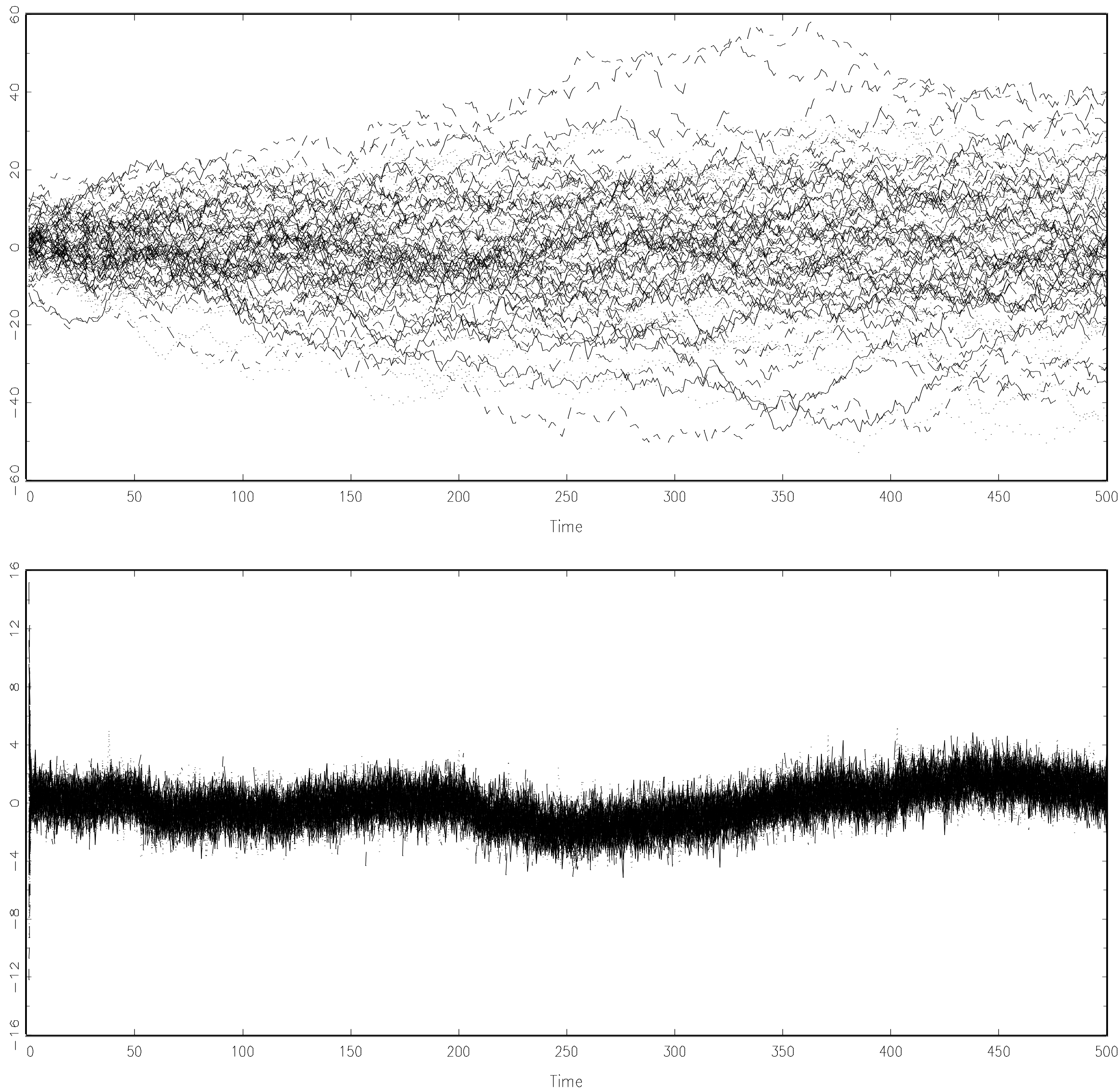


Figure 5: Results for the stochastic model using errors and factors with fat tails. First panel:$\rho = 0.9$, Second panel: $\rho = 0.999$