Commercial and Residential Mortgage Defaults: Spatial Dependence with Frailty*

Andrii Babii[†] UNC Chapel Hill
Xi Chen[‡] Freddie Mac
Eric Ghysels[§] UNC Chapel Hill

March 14, 2019

Abstract

We investigate the spatial dependence between commercial and residential mortgage defaults. A new class of observation-driven frailty factor models is introduced to do so. The idea of dynamic parameters embedded in the class of GAS models is utilized to estimate dynamic models of default risk with potentially multiple factors which are driven by stratified grouping of large panels of mortgage loan records. The score dynamics in the models is driven by so-called generalized residuals, and have therefore a fairly intuitive interpretation of ARMA-like dynamics. The asymptotic analysis recognizes the fact that we deal with both cross-sectional and time series data features. The proposed models are computationally easy to implement and therefore attractive in big data applications, something that gives them a considerable advantage in comparison to the typical latent factor frailty models proposed in the literature. Our empirical analysis demonstrates strong spatial dependence between commercial default and residential defaults.

[*]We greatly benefited from comments made by Jonathan Hill, three valuable referees, Frank Diebold and participants at the *Big Data in Predictive Dynamic Econometric Modeling* conference held at the University of Pennsylvania. The opinions expressed in this paper are those of the authors, and are not necessarily those of Freddie Mac or its Board of Directors.

[†]Department of Economics, UNC Chapel Hill, NC 27599, USA.

[‡]Quantitative Analytics, Senior. Freddie Mac, 1551 Park Run Dr, McLean, VA 22102.

[§]Department of Economics UNC, Department of Finance, Kenan-Flagler Business School, Chapel Hill, NC 27599, USA, LFin, Louvain-la-Neuve, Belgium and CEPR.

1 Introduction

One of the obvious challenges encountered in the econometric analysis of big data is coming up with computationally feasible methods with appealing statistical properties able to handle mega out-sized datasets. We consider the class of Generalized Autoregressive Score (GAS) models proposed by Creal, Koopman and Lucas (2013) and Harvey (2013) in a big data setting and study its statistical properties as well as its appealing features in a particular area of application involving big data. We study datasets consisting of millions of records pertaining to commercial mortgages in the U.S. retail industry and residential mortgage loans.[1] Modeling credit risk is an important area of interest to academics and practitioners. With the vast loan data, we try to tease out the complex cross-sectional and temporal heterogeneity across borrowers. In this respect, we propose a new approach which allows for an appealing degree of flexibility while keeping a fairly tight parametric model. Finally, it is important to note that while we study mortgage data, the class of credit models we propose applies to other types of contracts, such as credit card or student loans debt.

There are two parts to default risk: (a) idiosyncratic risk related to a specific loan which is usually tracked by the financial information pertaining to borrowers (e.g. borrowers' credit score, or debt-to-income ratio (DTI) which is one way lenders measure an individual's ability to manage monthly payment and repay debts) and (b) systematic risk which affects all mortgages and can be approximated by macro variables such as GDP, unemployment rates, etc. Frailty models start from the recognition that macro variables and mortgage-specific covariates may not be enough to explain all the risks. The main tenet of the class of frailty models is that we need some additional factor(s) beyond observable characteristics of borrowers (either firms holding corporate debt or households with mortgage, car, credit card or student loan debt), and the observable state of the economy. Frailty models can be classified into two types - using a characterization put forward by Cox (1981): parameter-driven models and observation-driven models. Both types of models have been used to predict credit risk. For example, parameter-driven models have been used to track the credit risk of corporate debts in Duffie et al. (2009) and Koopman and Lucas (2008), and to forecast mortgage default in Kau, Keenan and Li (2011). Meanwhile,

[1]The complete data set contains 1,683,703 records of commercial mortgages in the U.S. retail industry and the residential mortgage dataset consists of 1.09 billion records of single family mortgages. For the commercial loans, the records start in January 1999 and end in January 2016. There are 17,527 distinct mortgages and all of them are 10-year-balloon mortgages. The residential loans cover 24.2 million Fannie Mae distinct mortgages, all fully amortizing and fixed rate.

using observation-driven models, Creal, Koopman and Lucas (2013) and Creal et al. (2014) investigated corporate defaults.

Parameter-driven frailty models do not provide a computationally feasible solution in a big data context due to the lack of closed-form expressions for likelihood functions. Even the application of observation-driven frailty models has been confined to datasets of modest size typically used in the corporate bond literature. Our paper is to the best of our knowledge the first big data application of GAS-type observation driven frailty models. The first contribution of our paper, building on our earlier work, Chen, Ghysels and Telfeyan (2016), is that we propose a new class of spatial frailty models linking commercial and residential mortgage risk. We show how it is computationally fairly easy to implement GAS frailty models in a big data context. Moreover, we establish the asymptotic properties of the type of GAS models we consider, as they feature attributes not studied in the prior literature.

The high tractability of our frailty models plays a core part in exploring the causal relations of big datasets. Our computational tool is a fairly standard workstation with XEON E5 2680 CPU with 24 cores and a base frequency of 2.5 GHz.[2] Estimation typically takes about 15 minutes to half an hour even with datasets containing millions of records. For example, our largest residential dataset contains 52 million records. The models proposed in this paper, can be easily implemented for even larger datasets, but this would require implementation on clusters to speed up the computations. Such endeavor is beyond the scope of the current paper.

Our second contribution pertains to cross-market spillovers. Both in the academic literature and among industry applications it is common to separate commercial and residential mortgage risks. In academia, there is an extensive literature pertaining to residential mortgage design and risks, and there is a somewhat less voluminous literature treating commercial mortgages. In the financial industry, there is a similar organizational silos structure. Financial institutions usually have a residential real estate unit that is operating completely independently of the commercial mortgage department. In this paper, we break across these apparent barriers and study the spatial and temporal dependencies across two credit risk spheres. In reality, defaults in residential mortgages and commercial ones are interrelated. When a business initiates major layoffs, it will have ripple effects on the housing as well as retail sector in adjacent locations. Cities with major economic declines, such as Detroit, are prime examples revealing the interconnections between the default risks across households

[2]The workstation is shared and our estimation usually utilized 10 to 20 cores.

and businesses. Our empirical analysis demonstrates strong spatial dependence between commercial default and residential default in multiple respects. In particular, we apply Granger causality tests to the default rates of commercial mortgages and residential mortgages in the top 10 (Metropolitan Statistical Areas) MSA areas, and the test results reveal a significant lead and lag relationship for the two mortgage markets.[3]

The remainder of this paper is organized as follows. Section 2 describes the new class of models and the empirical implementation appears in Section 3. Conclusions appear in Section 4. Technical details with formal assumptions and proofs appear in the Appendix.

2 Spatial GAS frailty models

Reduced-form models assume that credit risk is not only related to the structural features of the household or firm but also depends on macro variables and financial information of borrowers. Research on reduced-form methods started with Beaver (1968) and Altman (1968) who used discriminant analysis as the main tool. Since Ohlson (1980) and Zmijewski (1984), binary response models like logit and probit regressions have been introduced into credit risk modeling. Most of these models estimate single-period default probabilities or credit scores. These reduced-form models assume that macro variables are able to explain all the systematic risk. However, Das et al. (2007) questioned the validity of this assumption and provided empirical evidence suggesting that macro variables do not fully account for systematic risk. Das et al. (2007), and subsequently Duffie et al. (2009), introduced the notion of common latent factors - also called frailty factors - which enter into the default intensity of proportional hazard models for credit risk. They showed that failure to control for these latent factors could cause downward biases in calculating Value-at-Risk. Since the seminal work of Das et al. (2007), increasing attention has been paid to credit risk modeling with frailty factors.

More directly related to the analysis in the current paper is Creal, Koopman and Lucas (2013) who designed a class of observation-driven models - called GAS models - and applied them to Moody's credit rating data. GAS models use scaled scores of

[3]For a definition of MSA, see https://en.wikipedia.org/wiki/Metropolitan_statistical_area.

the likelihood function to update the dynamics of latent factors. Using GAS models, Creal et al. (2014) jointly modeled macro variables and default outcomes with data of mixed-measurement and mixed-frequency. It is worth noting an appealing property of these models which makes their application to big data particularly attractive. Namely, when one applies the idea of a GAS specification to binary logistic regressions and chooses the intercept as the dynamic parameter, then the score of the dynamic intercept is actually the generalized residual defined in Gouriéroux et al. (1987) and the GAS models reduce to the generalized autoregressive moving average models proposed by Shephard (1995) and Benjamin, Rigby and Stasinopoulos (2003).

We are particularly interested in commercial and residential real estate loans and extend the work of Chen, Ghysels and Telfeyan (2016), who studied retail sector commercial loans using a simpler class of GAS frailty models compared to what is consider in this paper. In particular, we consider a new class of spatial credit risk models which exploit the fact that the location of mortgages is of special importance because the revenues of underlying properties largely depend on the local economy. Some prior work on the topic exists, but none is applicable to large loan datasets. Deng, Pavlov and Yang (2005) proposed a space-varying coefficient model for residential mortgages where the space-varying coefficients play the role of spatial latent factors. Their model involves a three-stage maximum likelihood estimation procedure and is therefore computationally too cumbersome to be implemented with large datasets. Non-parametric methods were also explored to capture the unobservable spatial variations of mortgage default. Using such methods, Follain, Ondrich and Sinha (1997) and Kau, Keenan and Li (2011) showed that MSA-level frailty has significant effects on mortgage termination risks. While their analysis is again confined to smaller datasets, it inspired us to adopt a similar MSA-level classification. Chen, Ghysels and Telfeyan (2016) did not study the asymptotic properties of the proposed model estimators. In this paper, we provide a rigorous econometrics analysis as well as a more comprehensive empirical application.

To set up the model, consider a set of n_t contracts at time t, and a vector y_t whose elements $y_{i,t}$, $i = 1, \ldots, n_t$, are binary and reflect the default status of loan i at time t. More specifically, the vector y_t consists of entries equal to one if a default happens at time t, and zero if the borrower can make timely payments. Note that the definition of a default event will be provided later in the empirical section, and we proceed here assuming it is well defined. Let us further denote the default (conditional) probability as $\pi_{i,t}$, i.e. $\pi_{i,t} = P(y_{i,t} = 1|I_{t-1})$, where I_{t-1} represents time $t-1$ information. To relate the default probability with the covariates, we adopt a binary logit model:

$$\pi_{i,t|t-1} = \exp(\mu_{i,t|t-1})/(1 + \exp(\mu_{i,t|t-1}))$$

where $\mu_{i,t|t-1}$ is a transformed default probability. It is defined as a linear combination of covariates and potentially multiple frailty factors to which mortgage i is exposed. To simplify notation we drop the conditioning information subscript $t-1$ and write:

$$\mu_{i,t} = \mu_{i,t}^0 + x_{i,t}^\top \beta$$

where the vector $x_{i,t}$ consists of a series of observed variables pertaining to time t but known the previous period $t-1$, such as financial information of the mortgage, β is a parameter vector common across all mortgages. Note that the covariates do not include a constant. Instead, the term $\mu_{i,t}^0$ represents the baseline default risk which is time-varying. A specific example would be

$$\mu_{i,t}^0 = c_i^g f_t^g + c_i^c f_t^c,$$

where f_t^g are geographic factors and f_t^c are credit score factors. Hence, in this example, the baseline default risk is driven by two types of (frailty) factors, one pertaining to spatial location and the other related to credit risk. For example, if we study d_g geographic regions then f_t^g is a $d_g \times 1$ dimensional vector process. Likewise, with d_c credit score factors, we have $d_c \times 1$ vector process. To illustrate this further, consider a geographic or spatial factor separates the data into 10 groups ($d_g = 10$), the top 10 MSA's, and the credit score factor involves 2 distinct groups, pooling contracts into prime and subprime ($d_c = 2$). In such a setting, let f_t^c represent the frailty factor related to the credit score with c_i^c, a vector of zeros except for a single entry equal to one which selects the credit score group to which loan i belongs, and f_t^g stands for the frailty factor pertaining to location with c_i^g the same type of selection vector with a single entry equal to one associated with the geographical location of the property. Hence, a loan for a property in Los Angeles assigns a unit entry to the MSA group covering LA county. Depending on whether the loan is prime or subprime will result in different selections via c_i^c.

The analysis in the remainder of this section is generic and therefore applies to residential as well as commercial real estate markets, or for that matter to any other type of credit risk setting. In a generic setting we assume there are K factor groupings:

$$\mu_{i,t}^0 = c_i^1 f_t^1 + \ldots + c_i^K f_t^K \equiv f_{i,t}, \tag{2.1}$$

hence, the model can be written as:

$$\pi_{i,t} = \exp(\mu_{i,t})/(1+\exp(\mu_{i,t})) \tag{2.2}$$

$$\mu_{i,t} = x_{i,t}^{\top}\beta + f_{i,t} \tag{2.3}$$

$$f_{i,t} = c_i^1 f_t^1 + \ldots + c_i^K f_t^K. \tag{2.4}$$

where $f_{i,t}$ is the time-varying intercept of the default intensity – a linear combination of the exposure of loan i to each of the K risk factors affecting the baseline default risk. It should also be noted that f_t^k is d_k-dimensional $k \in \{1, \ldots, K\}$. The vectors c_i^k are of dimension d_k for all i and k.

To specify $f_{i,t}$ as observation-driven we need to clarify a number of intermediate steps. First, we rely on an idea often explored in factor models, namely that individual observations can be classified into homogeneous groups - resulting in the notion of group factor models.[4] We assume loans bundled in the same group share the same frailty factors. This is the key assumption driving identification. In addition, we assume that c_i^k are predetermined for all i and k. This avoids ambiguity regarding the factor group classification of each individual loan. Hence, we do not address the more challenging issue of group allocation uncertainty sometimes considered in the group factor literature - see e.g. Ando and Bai (2016) and references therein. Since contracts in different groups have distinct factors, every grouping criterion corresponds to a separate set of factors. For example, as noted before, in the empirical application we consider groupings that are spatial. Hence, the spatial grouping stratifies the data according to the MSA where a property is located.

In what follows, we adopt the GAS-model paradigm and assume that the frailty factors have an autoregressive form. In particular, each of the K factors, with $\dim(f_t^k)$ $= d_k$, $k = 1, \ldots, K$, are stacked in a vector f_t of dimension $d = \sum_k d_k$ which satisfies the following recursion:

$$f_t = \delta + \theta f_{t-1} + \alpha s_{t-1}, \tag{2.5}$$

where the time subscript to the innovations s_{t-1} indicates that it involves using time

[4] See for example, Tucker (1958), Dauxois and Pousse (1975), Krzanowski (1979), Schott (1988), among others. Recently, Andreou et al. (2016) presented the asymptotic theory for model selection and estimation of group factor models with large N and T data.

$t-1$ information. In particular the vector s_{t-1} consists of stacked s_{t-1}^k where:

$$s_{t-1}^k \equiv \begin{bmatrix} s_{1,t-1}^k \\ \vdots \\ s_{d_k,t-1}^k \end{bmatrix} = \bar{y}_{t-1}^k - \hat{y}_{t-1}^k \equiv \begin{bmatrix} \bar{y}_{1,t-1}^k \\ \vdots \\ \bar{y}_{d_k,t-1}^k \end{bmatrix} - \begin{bmatrix} \hat{y}_{1,t-1}^k \\ \vdots \\ \hat{y}_{d_k,t-1}^k \end{bmatrix} \tag{2.6}$$

where each element j of the vectors on the RHS of the above equation satisfies:

$$\begin{aligned} \bar{y}_{j,t-1}^k &= \frac{1}{N_{j,t-1}^k} \sum_{i=1}^{n_t} y_{i,t-1} \mathbb{1}_{c_i^k = \iota_j^k} \\ \hat{y}_{j,t-1}^k &= \frac{1}{N_{j,t-1}^k} \sum_{i=1}^{n_t} \hat{\pi}_{i,t-1} \mathbb{1}_{c_i^k = \iota_j^k} \end{aligned} \tag{2.7}$$

where $N_{j,t-1}^k$ is the number of observations belonging to group $j \in \{1, \ldots, d_k\}$ of factor with $k \in \{1, \ldots, K\}$ and n_t is the total number of observations in period $t = 1, \ldots, T$. Finally, ι_j^k is a d_k-dimensional vector of zeros except of entry j equal to one.

The choice of innovation is crucial in updating the dynamics of the frailty factors. As pointed out by Chen, Ghysels and Telfeyan (2016), the score-driven dynamics for the intercept parameter in equation (2.3) amounts to using generalized residuals, as dubbed by Gouriéroux et al. (1987). In spatial applications this amounts to counting how many loans pertaining to properties located in MSA j at time $t-1$, $j = 1, \ldots,$ 10 (if we consider the top 10 MSA's). The equations in (2.7) imply that $\bar{y}_{j,t-1}^k$ is the empirical default rate for group j at time $t-1$, and $\hat{y}_{j,t-1}^k$ is the fitted default rate for the same group. The same logic applies to other factors, like for instance the FICO score which will track the defaults among loans within the same credit risk group. The innovation, therefore, measures the distance between models and data. If the innovation term is positive (negative), then the empirical PD is larger (smaller) than fitted one.[5]

[5] There is some prior work on GAS models involving time-varying spatial dynamics. Namely, Blasques et al. (2016) use the term spatial, although in a setting that is different from ours. Their model builds on the well-known spatial lag model for panel data. The strength of contemporaneous spillover effects is summarized in a single time-varying parameter, which is the spatial dependence parameter. In their application, they measure the time-varying cross-sectional dependence in European sovereign credit spreads and argue that the time-varying spatial dependence parameter may be interpreted as a measure of sovereign systemic risk.

One technical remark pertaining to the dynamics of factors is in order. It is standard in the GAS literature to scale the score function by its standard deviation. In our model, the GAS dynamics are driven by generalized residuals, which are between -1 and 1 and do not correspond to the exact score. For this reason, we deviate from the standard GAS literature and do not scale the dynamic equation.

The original class of GAS models proposed by Creal, Koopman and Lucas (2013) and Harvey (2013) is intimately related to the class of ARCH-type models, since the latter can be viewed as time-varying volatility parameter models. In the same vein, one could view the specification in equation (2.5) as reminiscent of multivariate GARCH models (see e.g. Bauwens, Laurent and Rombouts (2006) for further discussion). The analogy implies we also should be weary about parameter proliferation. Indeed, the specification in (2.5) involves matrices θ and α which are respectively a d-dimensional vector ($d = \sum_k d_k$, $d_k = \dim(f_t)$, $k = 1, \ldots, K$) and $d \times d$ matrix. Hence, for example with $d = 12$ we have 156 parameters. Restricted models will therefore be appealing. In subsection 3.2 we describe the empirical model specifications and elaborate on the restrictions we impose to avoid parameter proliferation. For the moment, we proceed with a generic specification.

The apparent similarity to multivariate ARCH-type models is a bit deceiving from a technical point of view. More specifically, the GAS version of binary choice models we study is non-trivial due to the fact that the frailty factors drive the dynamics of default risk in a discontinuous way. Our model is therefore similar to threshold models (see e.g. Tong (1983)), although the latter are simpler as they have a linear additive structure and are typically univariate. Our model does not admit such a simple form and it is, therefore, harder to deal with the discontinuities. We deal with these challenges in the Appendix containing the detailed theoretical results. Some of the key ingredients include invertibility in nonlinear time-series models – see the recent work by Wintenberger (2013) and Blasques et al. (2018). In particular, we distinguish the likelihood function initiated from any point in the parameter space versus the one drawn from the stationary distribution.

We simplify some arguments commonly used for ARCH-type (and GAS) models – see e.g. Francq and Zakoïan (2011) – which involves establishing conditions on approximate Hessians that are typically rather involved. More importantly, in terms of asymptotic normality, we cannot rely on the martingale difference sequence CLT commonly used in the GAS and GARCH literature. Instead, we use the CLT based on mixing conditions.

Before we state the main theoretical result, we should note that we are in principle dealing with asymptotic analysis along two dimensions, namely cross-sectionally as $N^k_{j,t-1}$ grows across all k and across time $t = 1, \ldots, T$. While there is now a well established asymptotic theory for both large cross-sections and time series in *linear* factor models, including group factor models - see Andreou et al. (2016), there are no results for GAS-type large panel data models of the type studied here. In fact, the asymptotic analysis of GAS-type models is mostly confined to univariate cases. Straumann and Mikosch (2006) establish the conditions for stationarity, ergodicity, and invertibility for general stochastic recurrence equations. Blasques, Koopman and Lucas (2014) and Blasques et al. (2018) exploit their results to establish root-T convergence and asymptotic normality of general univariate GAS models. The strategy in these papers consists of establishing the geometric ergodicity relying on the stochastic contracting Lipschitz properties, see also Bougerol (1993). Our model features more complex discontinuous dynamics, which is not Lipschitz and does not naturally fit into that framework. We use instead the theory of Markov chains, in particular, results of Feigin and Tweedie (1985). We are dealing with both multivariate models and large dimensional panels. We will treat the cross-section sampling as subordinated to the time series.

We define the log likelihood function for a generic K factor GAS model appearing in equations (2.3) through (2.7). We collect the parameters into a vector $\phi \equiv (\alpha, \beta, \delta, \theta)$, with possibly tight restrictions imposed on the general specification. Then the logistic quasi log likelihood and resulting QMLE estimator $\hat{\phi}_T$ can be written as:

$$\begin{aligned}\mathcal{L}_T(\phi) &= \frac{1}{Tn}\sum_{t=1}^{T}\sum_{i=1}^{n}\left[y_{it}\ln\left(\frac{\exp(\mu_{i,t})}{1+\exp(\mu_{i,t})}\right) + (1-y_{it})\ln\left(\frac{1}{1+\exp(\mu_{i,t})}\right)\right] \\ \hat{\phi}_T &\equiv \arg\max_{\phi\in\Phi}\mathcal{L}(\phi), \end{aligned} \tag{2.8}$$

where for simplicity we assume that the number of observations n is the same in all time periods. Maximizing the log likelihood results in an estimator $\hat{\phi}_T$ with the following asymptotic properties:

Theorem 1. *Under assumptions of Theorem A.2 as $T \to \infty$*

$$\sqrt{Tn}(\hat{\phi}_T - \phi_0) \xrightarrow{\mathrm{d}} N(0, J^{-1}VJ^{-1}),$$

where $V = \lim_{T\to\infty} \mathrm{Var}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\nabla l_t(\phi_0)\right)$.

We recognize that due to the QMLE nature of the estimator and the mixing CLT, we have to use Newey-West HAC estimators to compute the asymptotic covariance matrix of the parameter estimates. Note also that the asymptotic variance simplifies if the true distribution of conditional probabilities corresponds to the logistic distribution. In such a case

$$\sqrt{Tn}(\hat{\phi}_T - \phi_0) \xrightarrow{\text{d}} N(0, J^{-1}).$$

Note that it is preferable to estimate the robust standard errors based on the sandwich covariance matrix in order to be robust against the model misspecification. All the theoretical results are proven in Appendix sections A.1 (asymptotics QMLE) and A.2 (geometric ergodicity).

3 Empirical Results

We examine both commercial and residential mortgages and apply our new class of GAS models suitable for big data applications to each type of loans. The commercial mortgages dataset is provided by Trepp LLC, which is a leading provider of real estate data, and the residential mortgages dataset is published by Fannie Mae. The focus of our analysis is the top 10 MSA areas. Note that there are over 900 MSA areas, and a lot of them do not have sufficient data to generate reliable estimates of default probabilities. The top 10 MSA areas cover more than 30 percent of the observations for commercial loans in the entire dataset, which we felt struck a good compromise between having individual MSA's with sufficient loans and a substantial coverage of the entire database. Similarly, for residential mortgages, the original dataset contains 1.09 billion records of residential mortgages in the U.S. market. There are 24.2 million distinct mortgages and all are fixed-rate fully amortizing mortgages. We focus exclusively on mortgages in the top 10 MSA areas, which amounts to 0.28 billion records and 6.8 million distinct mortgages or approximately 28 percent of the original dataset published by Freddie Mac.

Using the aforementioned data, we estimate models pertaining to each credit sphere separately – models for commercial and residential markets. We also examine the interactions between the two markets via model specifications covering the joint probability of defaults across the two spheres.

3.1 Data

For commercial mortgages, the records start in January 1999 and end in January 2016. The origination of mortgages in our sample is between January 1995 and December 2015, and we focus exclusively on 10-year-balloon type loans. Table 1 provides detailed information about the top 10 MSA areas. Among these MSA areas, Los Angeles has the most number of mortgages, namely 1165, and the DC area has the least, namely 348.

For residential mortgages, the time period covered is from February 1999 to March 2016, and the origination time of the mortgages is between January 2000 and 2016. Table 2 provides further details, namely: the New York area has the most number of records (52 million) and the Los Angeles Area has the most number of loans (1.5 million). The Houston area has the least: 16 million records and 0.34 million loans. For the other regions, the number of records is generally around 20 to 30 million, and the count of mortgages is between 0.4 million and 0.8 million.

The default of residential mortgages is defined as 3 months or more delinquency, and loans with less than 3 months delinquency are considered current. The default of commercial mortgages is defined as 3 plus month delinquency, non-performing matured loans, loans in foreclosure and loans in REO status.[6]

3.2 Empirical Model Specifications

The purpose of this subsection is to describe the empirical models we fit to the data. The discussion is arranged as follows. First, we describe the covariates for respectively the commercial and residential real estate models. Next, we describe the model specifications for the frailty factor dynamics.

3.2.1 Covariates - Commercial Real Estate

The empirical model specifications for commercial mortgages are in part inspired by the empirical results reported in Chen, Ghysels and Telfeyan (2016). Here we

[6] We consider defaulted loans terminated if the loans feature one of the following status: non-performing matured, foreclosure or REO. Such loans are in serious delinquency and will most likely stay in default status once delinquent. This implies a default probability of one, regardless of all explanatory variables. Therefore, we remove the loans from the sample once they enter into a serious delinquency status.

augment their analysis with spatial frailty factors based on MSA groupings. In addition, we also expand the set of covariates beyond those used by Chen, Ghysels and Telfeyan (2016).

Recall that the default risk can be divided into an idiosyncratic and a systematic component. After extensive model specification tests, Chen, Ghysels and Telfeyan (2016) consider mortgage age and the debt service coverage ratio as the drivers of idiosyncratic risk. Mortgage age is defined as the number of months passed since the initiation of a mortgage loan. It has been widely used in the literature to explain the trend of mortgage default, which is usually approximated by a continuous function. The selection of this function largely depends on the amortization types of mortgages. Our commercial mortgages data are all 10-year-balloon mortgages and are therefore partially amortizing. In the case of partially-amortizing mortgages, usually, there are two default clusters observed during the life span of mortgages: one around the fifth year and the other at maturity. The former is due to deteriorating financial conditions of the borrowers, and the latter is caused by a failure to refinance. Since we found similar default clusterings in our data, we design a piecewise linear function to capture the influence of mortgage age on default rates. This function is called the age function and has the following form:

$$
\begin{aligned}
age_1 &= min(age, 56)\\
age_2 &= max(min(120-56, age-56), 0)\\
age_3 &= min(max(age-120, 0), 3)
\end{aligned}
$$

Resulting in the following age function:

$$Age\_function = age_1\beta_1 + age_2\beta_2 + age_3\beta_3$$

Debt Service Coverage Ratio (DSCR) is another important variable in modeling mortgage default. It is defined as the ratio of net operating income and current debt obligations. DSCR larger than one indicates borrowers have enough cash flow to make monthly payments. Values below one imply that borrowers may default. We use the original value of DSCR without any transformation. Models using these explanatory variables, together with a constant parameter for the intercept are called *static* in Table 3.

Apart from age function and DSCR, we consider a variety of variables related to property status and loan origination to capture the idiosyncratic risk of mortgages. For loan origination, we considered dummy variables derived from the loan purpose, namely purchase, refinance, or acquisition. Initially, we set up dummy variables for every type of loan purpose and found that among these loan purpose dummy variables, only the acquisition dummy was significant. In light of this, we decide to make the loan purpose variable a binary categorical variable and let it indicate whether the loan purpose is the acquisition or not. The variable is called 'Acquisition Flag', with value 1 for loan purpose as the acquisition.

To explain idiosyncratic risk related to property status, we created two series of variables: one describing the anchor types of the properties and the other related to the year the properties were built. The properties considered in this paper are all retail properties and it is well known that anchor types strongly influence the profitability of retail properties. For example, a thriving shopping mall usually has large retail stores inside and the large retail stores are the 'anchors' of the mall. The anchor stores attract customers and bring cash flows for all stores in the shopping mall. Therefore, the anchor types of retail properties have a strong impact on the default status of the mortgages and consequently, we use the anchor type as a categorical variable in the model. Specifically, we derive three dummy variables from the anchor type and call them 'anchor flag', 'shadow anchor flag', and 'single tenant flag'. These flag variables describe whether the underlying properties for mortgages are anchored, shadow anchored or occupied by a single tenant.[7] Beside using anchor types in the model, we also derive two variables related to the property year: one is property year and it records when the property was built; the other one is property age and it is the difference between the property built-year and the mortgage origination year.

The variables discussed above all relate to a specific mortgage and explain idiosyncratic risk only. For systematic risk, we use variables related to MSA-specific economic conditions. Specifically, we use the Commercial Property Price Index (CPPI) published by Real Capital Analytics (RCA), a leading data provider for the commercial real estate, which measures price appreciations of commercial properties at the metro level.[8] Based on CPPI, we construct 3 variables which measure systematic risk related to house price fluctuation:

- current CPPI without any transformation. It reflects the systematic risk caused

[7]There are other possible anchor types, but our empirical analysis has shown that these anchor types are either rare or not significant in modeling.

[8]RCA CPPI covers industries of retail, office and industrial. While the coverage of price index is broader than retail properties studied in this paper, the common trends among these three industries are identical. Therefore, the bias of using RCA CPPI is relatively minimal and the details are discussed in the paper.

by house price fluctuation. High CPPI generally implies low systematic risk.

- changes in CPPI since the origination of mortgages, defined as the ratio between current CPPI and initial CPPI at origination time. This variable supplements the current CPPI by measuring the systematic risk relative to a specific mortgage, since it incorporates the macro environment at origination. Accordingly, a large negative change indicates high systematic risk.

- current LTV, which is defined as the ratio between the current Unpaid Principal Balance (UPB) and current property value. The current property value is calculated by scaling the initial property value with the CPPI change since origination. This variable further differentiates the systematic risk embedded in a mortgage with respect to different LTV.

These three variables capture systematic risk on different levels and arguably provide a comprehensive characterization of systematic risk for commercial mortgages.

So called *dynamic* models, appearing in Table 3, include spatial frailty to capture systematic risk besides the aforementioned variables.

3.2.2 Covariates - Residential Real Estate

The model specifications for residential mortgage loans appear in Table 4. We consider a large set of variables to model idiosyncratic risk of residential real estate mortgages. They cover various loan-specific risk factors pertaining to borrowers, mortgage features, properties, and sellers of the mortgages.[9] There are a total of 14 variables used in analysis and we describe the details in the following paragraphs.

Variables related to borrowers. Borrowers' financial strengths and credit history strongly influence the underlying risk of mortgages. There are five such variables in our models: Debt to income ratio, flag for first time home buyer, number of borrowers, subprime borrowers, and co-borrower credit scores. Debt to income ratio is extensively used in the literature to model the idiosyncratic risk of residential mortgage borrowers. It is defined as the ratio of monthly loan payment and the borrower's income. A DTI larger than one implies that the borrower cannot cover the monthly

[9]The loans we consider here are all securitized by Fannie Mae, and they are purchased from sellers of mortgages across the US. Note that the sellers of mortgages may not be the originators.

payment with her/his income. DTI measures the financial capacity of borrowers and is similar to DSCR for commercial mortgages. A notable difference between DTI and DSCR is the frequency of data availability: DTI is only measured at the origination time of mortgages, and DSCR is updated as it is based on the most recent financial statements of borrowers. Therefore, DTI is a static variable and actually works as a fixed intercept. By contrast, DSCR is a dynamic variable and reflects the current financial status of borrowers.

The flag for first time home buyer is a dummy variable, with one indicating a first time home buyer and zero otherwise. Potentially, a first time home buyer may have a higher default risk than other borrowers, because they may have less savings. The number of borrowers for a mortgage ranges from 1 to 10. While the values have a large range, 99 percent of the mortgages have only 1 or 2 borrowers. More borrowers means greater risk sharing.

Two variables relate to credit scores of borrowers and they asses the credit history of borrowers. The first is an indicator for subprime borrowers and captures the difference between the credit quality of prime and subprime borrowers. Co-borrower scores report the credit capability of co-borrowers in the case of co-signed loans. A high co-borrower score means a low risk because if one borrower suffers financial hardship, the co-borrower with strong financial credentials could continue to make mortgage payments.

Variables related to loans. They are loan purpose, mortgage insurance percentage, and mortgage insurance type. Loan purpose is a dummy variable, zero if the borrower uses the loan to buy a house, and one if the loan is for the purpose of refinancing. Mortgage insurance percentage records the fraction of insurance borrowers purchased for the mortgages. Mortgage insurance is generally required by lenders if borrowers cannot make the 20 percent down payment. This variable is continuous and ranges from 0 to 100 percent. When mortgage insurance is required for low-down payment loans, they could be purchased by either borrowers or lenders, and about 20 percent of the mortgage insurance is paid by the lenders. So this variable enters into the model as a dummy variable, and value 1 indicates lender-purchased mortgage insurance.

Variables related to properties. They influence the borrowers' decision to default and three variables are used in the models: occupancy status, property types, and the number of units. Occupancy status reflects how borrowers' use the house and has three possible values: occupied, second home, and investment. Due to the three categories, this variable is modeled as two dummy variables. One dummy variable indicates whether the house is a second home or not and another one determines if the house is an investment. Accordingly, "occupied" serves as a base case and is absorbed by the intercept.

The values of property types could be single-family, condo, co-op, manufactured housing or planned unit development (PUD). Among all these types, single-family and condo make up most of the population. This variable enters into the model as a categorical variable and single-family works as the base category. The number of units in one property ranges from 1 to 4 and properties with more than 5 units are considered as multi-family or apartment by Fannie Mae. About 95 percent of the properties are single-family, and this variable works as a dummy variable, with value 1 indicating number of units are more than 1.

Variables related to sellers. Channel is a variable reporting from whom Fannie Mae purchased the loan from and the values are retail, broker, or correspondent. Retail denotes loans purchased from banks originating the mortgages directly, and the other two values indicate sellers who do not originate loans directly. The distribution of these three channels is roughly even, although mortgages from brokers occur less often than the other two. This variable is categorical and retail works as the base category. Finally, analogously to the commercial real estate models, the specifications with all aforementioned variables and a constant parameter for the intercept are called *static* in Table 4.

Systematic risk variables. Similar to commercial mortgages, we use local house price fluctuations to capture the systematic risk of residential mortgages. The residential house price indices (HPI) are published by Federal Housing Finance Agency (FHFA).[10] The FHFA HPI is, compared to the Case-Shiller index, more appropriate for our analysis in terms of coverage, as it only uses transaction data of mortgages securitized by Fannie Mae and Freddie Mac. Since our residential data are provided

[10] FHFA is the regulator for Fannie Mae and Freddie Mac, collects their data and regularly publishes information about housing markets.

by Fannie Mae, the FHFA HPI, therefore, better reflects the movement of house prices for the loans in our sample. Using FHFA HPI, we construct three variables for residential models, including current HPI, HPI change since origination, and the current LTV. The definitions of these variables are similar to those used in the commercial real estate mortgage models.

3.2.3 Frailty factor dynamics

Given the high-dimensional nature of the factors, a major concern is parameter proliferation. We consider various specifications. Let us recall equation (2.5), namely:

$$f_t = \delta + \theta f_{t-1} + \alpha s_{t-1},$$

where each of the K factors, with $\dim(f_t^k) = d_k$, $k = 1, \ldots, K$, are stacked in a vector f_t of dimension d ($d = \sum_k d_k$) and the innovations s_{t-1} consists of stacked s_{t-1}^k as specified in equations (2.6)-(2.7).

We consider two types of specifications in our empirical research. They are:

Unconstrained small K We start with the most general model but refrain from using a lot of factors. This unconstrained specification will allow us to study interactions between commercial and residential real estate default risk for a *given* MSA. This means that $K = 2$, as there will be a frailty factor specific to commercial defaults and one specific to residential ones. The model is estimated for each MSA separately and is therefore specified as:

$$\begin{bmatrix} f_t^r \\ f_t^c \end{bmatrix} = \begin{bmatrix} \delta_r \\ \delta_c \end{bmatrix} + \begin{bmatrix} \theta_{rr} & \theta_{rc} \\ \theta_{cr} & \theta_{cc} \end{bmatrix} \begin{bmatrix} f_{t-1}^r \\ f_{t-1}^c \end{bmatrix} + \begin{bmatrix} \alpha_{rr} & 0 \\ 0 & \alpha_{cc} \end{bmatrix} \begin{bmatrix} s_{t-1}^r \\ s_{t-1}^c \end{bmatrix}, \tag{3.9}$$

$$\begin{bmatrix} s_{t-1}^r \\ s_{t-1}^c \end{bmatrix} = \begin{bmatrix} 1/N_{r,t-1} \sum_{i:c_i^r=1} y_{i,t-1} \\ 1/N_{c,t-1} \sum_{i:c_i^c=1} y_{i,t-1} \end{bmatrix} - \begin{bmatrix} 1/N_{r,t-1} \sum_{i:c_i^r=1} \hat{\pi}_{i,t-1} \\ 1/N_{c,t-1} \sum_{i:c_i^c=1} \hat{\pi}_{i,t-1} \end{bmatrix}$$

where $N_{r,t-1}$ and $N_{c,t-1}$ are the number of observations for respectively residential and commercial loans at time $t-1$ for a given MSA (since the models are estimated separately for each individual MSA).
Note that the parameters θ_{rc} and θ_{cr} in the above model allow us to test for Granger causal patterns between residential and commercial real estate defaults and vice versa. We estimate the model for each of the 10 MSA's separately and therefore ignore the common beta restriction. We also experimented with estimating all MSA's

jointly, i.e. imposing the common exposure to DSCR and the age function and all other aforementioned covariates. We found that having a separate β yielded better results, and these will be reported.

Diagonal GAS We label the next specification as diagonal GAS, similar to diagonal GARCH, where each of the K factors, with $\dim(f_t^k) = d_k$, $k = 1, \ldots, K$, satisfy the following recursion:

$$f_t^k = \delta_k + \text{diag}(\theta_k^1, \ldots, \theta_k^{d_k}) f_{t-1}^k + \text{diag}(\alpha_k^1, \ldots, \alpha_k^{d_k}) s_{t-1}^k. \tag{3.10}$$

We estimate this type of model for residential and commercial real estate separately. Hence, the focus is no longer on the relationship between the two credit spheres. Instead, the emphasis is on studying either the residential or the commercial real estate market. In the unrestricted model, there may be Granger causal patterns between frailty factors, whereas in the diagonal GAS all such lead-lag patterns are shut down and we are dealing with a stacked system of univariate GAS models. This has implications not only in terms of model specification but also in terms of estimation as the univariate models can be treated separately (although they still share a common β parameter vector). Ignoring all Granger causal patterns will be an appealing approach for large K, because of its computational simplicity with big data, as one can always examine such patterns in subsequent tests.[11]

We adopt the diagonal GAS specification appearing in equation (3.10) with $d_1 = 10$, the number of MSA's we consider. Recall that this is essentially a stacked system of 10 univariate GAS models, each having their own recursive score equation parameters and β as a common parameter.

3.3 GAS Model Estimation Results

We discuss first models estimated for each MSA separately – that is diagonal GAS models – followed by models involving joint estimation. Regarding the former, due to the large volume of empirical results, we cover only two MSA's, namely LA and NYC in the main body of the paper while more detailed results appear in the Online

[11]Pierce and Haugh (1977) advocated a similar approach in the context of ARIMA models. Although it has been argued in the context of such models that there are many Granger causality tests with better properties, we find their approach appealing in a big data environment as the diagonal GAS model specification vastly simplifies estimation.

Appendix. We pick LA and NYC as they are the top 2 MSA's in both the commercial and the residential loan market. Before presenting joint estimation results, we report evidence about Granger causality between residential and commercial default risks.

3.3.1 Individual MSA estimation results - Residential Real Estate

Dynamic and static model estimation results pertaining to residential mortgages for LA and NYC are reported in respectively Tables 5 and 6. Parameter estimates and HAC standard errors are reported for each of the tables.

A set of standard covariates related to borrowers, namely DTI, Flag of first time home buyer, Subprime borrower, Loan purpose (non-purchase), Occupancy status (second home), and Occupancy status (investment) feature in the static I and dynamic I models. They are all significant, using HAC standard errors, and the dynamic model has a better fit according to the likelihood, AIC, and BIC. It is worth noting that the results for LA show a much more dramatic improvement when the frailty factor is added. Adding loan characteristics as covariates – as in the Static and Dynamic models II – results in further improvements in terms of likelihood and the two information criteria. Variables related to HPI will be quite important in that regards, as will be discussed shortly. The third model specification adds variables related to sellers and systematic risk measures. Again, the model improves along the same lines. The Dynamic class of models, involving a (spatial) frailty factor always dominate their static counterpart and ultimately, Dynamic model III is the best in-sample fit.

A few comments regarding the parameter estimates of the covariates are in order. Starting with DTI, we find extremely large estimates with Static I, while the dynamic models yield positive significant parameters that are smaller by more than a factor of ten. More interesting to note is that the flag for first time home buyer yields positive estimates for Static models, hence increasing default risk, whereas the Dynamic models have significant negative estimates - thus implying reduced risk. The results in the Online Appendix indicate that this result applies to the MSA's beyond the two covered here. All specifications agree on the fact that subprime borrowers have increased default risk, although it is noteworthy that estimates for the Dynamic IV are much larger. The number of borrowers has the wrong sign in Static III model, while it is negative and significant for Dynamic III and IV models, as expected. Current LTV and HPI change yield similar parameter estimates across models. In contrast, current HPI yields insignificant estimates in Dynamic models. Co-ops have reduced risk - although not in Static models, whereas manufacturing housing does not seem to have an effect on default risk. Planned unit developments (PUD) have a small significant impact, again only in Dynamic models. A sign reversal also appears with the condo dummy. All model specifications agree that the channel from whom Fannie Mae purchased the loan does not matter. We should expect that the co-borrower FICO score lowers default risk and that is indeed the case for Dynamic models, but again not of Static ones. Borrowers insurance does not impact default risk according to the Dynamic model estimates. Count of borrowers should reduce risk, it does in Dynamic models, but not in static ones. Overall, we can conclude that in many cases the point estimates of Dynamic models have the expected signs when significant, and that is not the case with the Static counterparts. Hence, adding frailty factors not only improve the fit, but they also seem to align the impact of covariates.

According to the theoretical results covered in the previous section, we know that stationarity for the dynamic class of models requires in the single factor case that $\theta + \alpha/4 < 1$ (Assumption A.1.3). That requirement is easily satisfied for all the dynamic models estimated with the LA MSA data. For NYC this is the borderline case for Dynamic model I, with only a limit set of covariates pertaining to the borrower characteristics. Since this model is misspecified it should not be a major concern.

In the Online Appendix we report for the other eight MSA's and we find results that are broadly in line with the findings appearing in Tables 5 and 6. Dynamic model III, that is the model with all covariates and spatial frailty factor, is always the best model. In two cases, the Dallas and Houston MSA's, we note that the point estimates of the frailty factor dynamics violate the stationarity conditions for Dynamic model III.

The likelihood function and the information criteria feature major improvements when we turn to the last column pertaining to Dynamic model IV. Recall from Table 4 that this involves two frailty factors, namely a spatial as well as a credit risk frailty factor based on the FICO score of the borrower. The substantial improvement in the model fit indicates that a credit score frailty factor is clearly called for. This is not surprising, as the specification separates prime and subprime borrowers, both featuring different risk characteristics, as is widely reported in the literature.

To shed further light on the comparison of the models, we turn our attention to Figure 1 where we plot the empirical default probabilities alongside those obtained from the empirical models for NYC, DC, Dallas and Phoenix: two Eastern metropolitan areas and two Southern ones. While DC and Phoenix have high default rates, NYC and

Dallas have relatively low ones. Across all four metropolitan areas, we note that Fannie Mae mortgages had a peak in 2003-2004 followed by a smaller one during the recent financial crisis. The three static models completely miss those peaks. Static model I yields for all practical purposes a flat line, whereas static models II and III are virtually identical as the predicted default lines are almost indistinguishable. In contrast, dynamic models very much track to empirical default probabilities. There appears to be a slightly delayed pattern between empirical defaults and predicted ones, as one might expect from the GAS model formulation.

3.3.2 Individual MSA estimation results - Commercial Real Estate

According to the summary statistics reported in Tables 1 and 2 we note that the mortgage counts for LA are 1,582,224 residential versus 1,165 commercial and for NYC respectively 1,131,311 versus 842. Hence, it is fair to say that residential mortgage loans involve a large sample, whereas the commercial sample is small. This observation is even more relevant once we consider the less populated MSA's such as DC with only 348 contracts for commercial and 639,027 residential loans. Therefore we have to bear in mind that estimation of commercial loan default risk is more challenging.

Dynamic and static models estimation results for commercial mortgages for LA and NYC are reported in respectively Tables 7 and 8. Parameter estimates and HAC estimator standard errors are reported for each of the tables. The static I and dynamic I models only contain a restricted set of covariates, namely DSCR and age function. They feature the lowest likelihood function among respectively static and dynamic models. Indeed, adding the Current HPI, HPI change and Current LTV to the models vastly improves the fit as all three covariates are significant (with some minor exceptions for Current HPI). Recall that these three variables characterize systematic risk for commercial mortgages. Changes in HPI lower default risk, as expected, and higher LTV increases default risk, also as expected. The evidence for the next batch of covariates is more mixed, namely Anchor Flag, Shadow Anchor Flag, Single Tenant Flag, Acquisition Flag, Property Year and Property Age. Yet, based on AIC and BIC for static and dynamic models, the larger models are preferred, i.e. static III and dynamic III are the best models within their class. It is interesting to note that the parameter estimates for the covariates do not differ so much between Static and Dynamic models, in contrast to the findings with residential mortgages.

What about the frailty factor, i.e. the comparison of the class of static and dynamic models? The presence of spatial frailty clearly improves the fit, as we observed with residential loan models. We do observe a different pattern of θ and α parameters, however. While in the case of residential mortgages the former yielded estimates in the 0.85 to 0.95 range we now observe much lower estimates of the autoregressive parameter θ, with values around 0.75 whereas the values of α increased substantially. The stationarity conditions still hold, but it is clear from the parameter estimates that the frailty factors respond more directly to residuals, i.e. to past prediction errors. Allowing for a dynamic spatial frailty factor tends to increase the response of default risks to DSCR, as shown in the first lines of Tables 7 and 8.

3.3.3 Granger causality patterns between Commercial/Residential Defaults

We noted that in the interest of parsimony, we start with diagonal-type restrictions on the dynamics of the score process. Ignoring all Granger causal patterns is reminiscent of the approach suggested by Pierce and Haugh (1977). Note that we have various patterns to investigate. Namely, we can examine across MSA's for a particular credit market - commercial or residential. This would amount to studying the Granger causality between frailty factors across MSA's. Another causality pattern of interest would be between commercial and residential mortgages for a given MSA.

Tables 9 reports empirical results for Granger causal relationships among spatial frailty factors for commercial mortgages. The table reports the p-values of Granger causality tests between the frailty of commercial mortgages in main MSA areas. The length of lags included in causality tests is 6 months. The first column contains the dependent variables in the causality tests, and the first row contains the independent variable in the causality tests. The last row named 'Significant Leads' counts the number of p-values less than 0.05 in each column, and the last row named 'Significant Lags' counts the number of p-values less than 0.05 in each row. The results show that there is no significant lagged relationship for NYC, except for Miami. Conversely, NYC Granger causes all other MSA's except LA. The other MSA's show a more mixed pattern, with typically about half of the MSA's having some impact. It is worth noting that geographical adjacency is not necessarily a determinant of significant lead/lag patterns. For example, DC does not affect Philadelphia at 5% significance level, and vice versa.[12]

[12] In the Online Appendix we also report evidence regarding the lead/lag relationships among frailty factors for prime mortgages. Overall, there are strong Granger causal patterns across all MSA's.

Table 10 covers lead/lag patterns between prime residential and commercial frailty. More specifically, the entries to the table are p-values for Granger causality tests between prime residential and commercial loan defaults for two samples: (1) the entire sample and (2) a sample only covering the pre-crisis era. Granger causality is examined at one-, three-, six- and twelve-month horizons. It is interesting to learn from this table that while there does not seem much of any lead-lag pattern for the entire sample, except for the Chicago MSA, it is clear that prior to the subprime mortgage crisis there were some interesting lead-lag relationship between the residential market (the same applies to the subprime results not reported here for brevity) and the commercial loan market. More specifically, there is evidence of Granger causality from the residential to commercial real estate in Chicago, Houston, LA, Philadelphia, Phoenix, DC and to a lesser extent Miami (only at a 12-month horizon).

3.3.4 Joint Commercial/Residential Default Risk Model Estimates

The evidence reported in Table 10 prompts us to estimate joint models of frailty of the type presented in equation (3.9). Since our focus is on causal patterns running from the residential to the commercial market, we focus on a constrained version, namely:

$$\begin{bmatrix} f_t^r \\ f_t^c \end{bmatrix} = \begin{bmatrix} \delta_r \\ \delta_c \end{bmatrix} + \begin{bmatrix} \theta_{rr} & 0 \\ \theta_{cr} & \theta_{cc} \end{bmatrix} \begin{bmatrix} f_{t-1}^r \\ f_{t-1}^c \end{bmatrix} + \begin{bmatrix} \alpha_{rr} & 0 \\ 0 & \alpha_{cc} \end{bmatrix} \begin{bmatrix} s_{t-1}^r \\ s_{t-1}^c \end{bmatrix}, \tag{3.11}$$

Hence, in the above specification, the frailty of the residential market feeds into the that of the commercial market and therefore mimicking the potential causal pattern documented in Table 10. We estimate the above model for each MSA. Therefore we have per MSA a joint model of frailty interactions between the residential and commercial market with emphasis on the one-directional impact of the former on the latter. In Table 11 we report the estimation results for the 10 MSA's. Regarding the covariates, we estimate Dynamic model III specifications for both the residential and commercial markets. The parameter estimates for the covariates pertaining to the residential default risk are pretty much the same as in those reported in respectively Tables 5 and 6 for LA and NYC and the Online Appendix for the other MSA's. Hence, we only report the full set of parameter estimates for the commercial side of the joint model. While there are some small differences in the parameter estimates for the covariates, we find that overall the results are quite similar. Hence, the joint or individual MSA estimates of the covariate parameter estimates are largely unaffected by the estimation procedure being used. Our focus is therefore on the frailty dynamics. Instead of three key parameters, we now have four, including θ_{cr}, which is the key parameter of interest here.

A word of caution is in order regarding the comparison between the Granger causality test results in Table 10 and the empirical model estimates in Table 11, particularly the parameter estimates of θ_{cr}. It is important to note that the settings of the causality tests and those for the joint models are not exactly the same. The key difference is that the score function in the model dynamics features as a moving average term. As our models satisfy the invertibility conditions, the MA term can be written as an infinite order AR component. Meanwhile, the causality tests in Table 10 are based on finite order AR. We, therefore, may expect differences between the two findings, and indeed we do. Turning our attention to the estimation results for θ_{cr} reported in Table 11, we find significant and positive values for Atlanta, Houston, and Phoenix. Chicago could be added if we use a one-sided test that the parameter is positive and significant. This means there is some overlap with the Granger causality results reported via simple regression models.

3.3.5 Stress testing

Stress testing is a standard practice for risk management in the banking system, and banks are required by regulators to evaluate asset values under stress scenarios. Such scenarios pertain to pre-specified macroeconomic conditions, such as drops in GDP or house price indices. In this process, no frailty risk is considered and systematic risk is assumed to be explained. In light of this, GAS models provide a new option to enhance the current practice of stress testing: banks can estimate the models of mortgage default with frailty and then simulate the default risk in different frailty scenarios. Different from the current method of stress testing, the frailty factors in the GAS models will capture the risk ignored by the current framework, and yield more accurate estimates of risk measures in an adverse environment.

In addition to loading unexplained systematic risk, the frailty factors can also work as a proxy for uncaptured model risk, which is typically ignored in the current framework of stress testing. As discussed before, the innovation terms in the frailty factor are the difference between empirical and predicted PDs, and include both unexplained systematic risk and forecasting errors. The latter are caused by model risk, and a GAS model with frailty factors offers an easy way to capture the model risk in stress testing. In addition, the parsimonious parameterization of GAS models enables efficient simulation for stress testing. To illustrate the use of GAS models for stress testing, we simulate a pool of 10,000 mortgages for 10 years at monthly frequency, and consider two scenarios: one is normal with low PD and could be used to calculate expected default costs; the other is stressed with high PD, and could be used to compute stressed default costs. The simulated PDs of the pool is represented in Figure 2, with the right panel as normal and left as stress. The simulation process is relatively fast and takes only seconds on a laptop.

4 Conclusion

A new class of observation-driven frailty factor models is introduced. The idea of dynamic parameters embedded in the class of GAS models is utilized to estimate dynamic models of default risk with potentially multiple factors which are driven by the stratified grouping of large panels of mortgage loan records. We examine in particular spatial MSA and FICO score groupings. We derive the asymptotic properties of the parameter estimators for generic GAS frailty models. The score dynamics in the models are driven by so-called generalized residuals and have therefore a fairly intuitive interpretation of ARMA-like dynamics. The proposed models are computationally easy to implement and therefore attractive in big data applications, something that gives them a considerable advantage in comparison to the typical latent factor frailty models proposed in the literature. In terms of application, we do find considerable in-sample improvements resulting from including frailty factors into otherwise standard (static) default risk models. On the downside, we also find that the data features non-trivial Granger causality patterns between different regions of the US both for the commercial and residential market segments. Capturing such patterns remains a challenging in terms of parsimonious model structures. To the best of our knowledge, we are the first to investigate the spatial dependence between commercial and residential mortgage defaults and to propose econometric methods applicable to big data environments. Yet, there is still considerable room for improvement in terms of joint dynamic specifications which we leave for future research.

References

Altman, Edward I. 1968. "Financial ratios, discriminant analysis and the prediction of corporate bankruptcy." *Journal of Finance*, 23: 589–609.

Ando, Tomohiro, and Jushan Bai. 2016. "Panel Data Models with Grouped Factor Structure Under Unknown Group Membership." *Journal of Applied Econometrics*, 31: 163–191.

Andreou, Elena, Patrick Gagliardini, Eric Ghysels, and Mirco Rubin. 2016. "Is Industrial Production Still the Dominant Factor for the US Economy?" Available at `SSRN:https://ssrn.com/abstract=2731334`.

Bauwens, Luc, Sébastien Laurent, and Jeroen VK Rombouts. 2006. "Multivariate GARCH models: a survey." *Journal of Applied Econometrics*, 21: 79–109.

Beaver, William H. 1968. "Alternative accounting measures as predictors of failure." *Accounting Review*, 43: 113–122.

Benjamin, Michael, Robert Rigby, and Mikis Stasinopoulos. 2003. "Generalized autoregressive moving average models." *Journal of the American Statistical Association*, 98: 214–223.

Blasques, Francisco, Paolo Gorgi, Siem Jan Koopman, and Olivier Wintenberger. 2018. "Feasible invertibility conditions and maximum likelihood estimation for observation-driven models." *Electronic Journal of Statistics*, 12: 1019–1052.

Blasques, Francisco, Siem Jan Koopman, and André Lucas. 2014. "Maximum likelihood estimation for generalized autoregressive score models." Tinbergen Institute Discussion Paper 14-029/III.

Blasques, Francisco, Siem Jan Koopman, Andre Lucas, and Julia Schaumburg. 2016. "Spillover dynamics for systemic risk measurement using spatial financial time series models." *Journal of Econometrics*, 195: 211–223.

Bougerol, Philippe. 1993. "Kalman filtering with random coefficients and contractions." *SIAM Journal on Control and Optimization*, 31: 942–959.

Chen, Xi, Eric Ghysels, and Roland Telfeyan. 2016. "Frailty Models for Commercial Mortgages." *Journal of Fixed Income*, 26: 16–31.

Cox, David R. 1981. "Statistical analysis of time series: some recent developments [with discussion and reply]." *Scandinavian Journal of Statistics*, 8: 93–115.

Creal, Drew, Bernd Schwaab, Siem Jan Koopman, and André Lucas. 2014. "Observation-Driven mixed-measurement dynamic factor models with an application to credit risk." *Review of Economics and Statistics*, 96: 898–915.

Creal, Drew, Siem Jan Koopman, and André Lucas. 2013. "Generalized autoregressive score models with applications." *Journal of Applied Econometrics*, 28(5): 777–795.

Das, Sanjiv R, Darrell Duffie, Nikunj Kapadia, and Leandro Saita. 2007. "Common failings: How corporate defaults are correlated." *Journal of Finance*, 62: 93–117.

Dauxois, J., and A. Pousse. 1975. "Une Extension de l'Analyse Canonique. Quelques Applications." *Annales Institut Henri Poincaré*, XI: 355–379.

Davydov, Yurii Aleksandrovich. 1973. "Mixing conditions for Markov chains." *Theory of Probability and its Applications*, 18: 312–328.

Deng, Yongheng, Andrey D Pavlov, and Lihong Yang. 2005. "Spatial heterogeneity in mortgage terminations by refinance, sale and default." *Real Estate Economics*, 33: 739–764.

Duffie, Darrell, Andreas Eckner, Guillaume Horel, and Leandro Saita. 2009. "Frailty correlated default." *Journal of Finance*, 64: 2089–2123.

Feigin, Paul D, and Richard L Tweedie. 1985. "Random coefficient autoregressive processes: a Markov chain analysis of stationarity and finiteness of moments." *Journal of time series analysis*, 6: 1–14.

Follain, James R, Jan Ondrich, and Gyan P Sinha. 1997. "Ruthless prepayment? Evidence from multifamily mortgages." *Journal of Urban Economics*, 41: 78–101.

Francq, Christian, and Jean-Michel Zakoïan. 2011. *GARCH models: structure, statistical inference and financial applications.* John Wiley & Sons.

Gouriéroux, Christian, Alain Monfort, Eric Renault, and Alain Trognon. 1987. "Generalised residuals." *Journal of Econometrics*, 34: 5–32.

Harvey, Andrew C. 2013. *Dynamic models for volatility and heavy tails: with applications to financial and economic time series.* Cambridge University Press.

Kau, James B, Donald C Keenan, and Xiaowei Li. 2011. "An analysis of mortgage termination risks: a shared frailty approach with MSA-level random effects." *Journal of Real Estate Finance and Economics*, 42: 51–67.

Koopman, Siem Jan, and André Lucas. 2008. "A non-Gaussian panel time series model for estimating and decomposing default risk." *Journal of Business and Economic Statistics*, 26(4): 510–525.

Krzanowski, WJ. 1979. "Between-groups Comparison of Principal Components." *Journal of the American Statistical Association*, 74: 703–707.

Meyn, Sean P, and Richard L Tweedie. 2012. *Markov chains and stochastic stability.* Springer Science & Business Media.

Ohlson, James A. 1980. "Financial ratios and the probabilistic prediction of bankruptcy." *Journal of Accounting Research*, 18: 109–131.

Pierce, David A, and Larry D Haugh. 1977. "Causality in temporal systems: Characterization and a survey." *Journal of Econometrics*, 5: 265–293.

Rio, Emmanuel. 2017. *Asymptotic theory of weakly dependent random processes.* Springer.

Schott, J. R. 1988. "Common Principal Component Subspaces in Two Groups." *Biometrika*, 75: 229–236.

Shephard, Neil. 1995. "Generalized linear autoregressions." DP Nuffield College - University of Oxford.

Straumann, Daniel, and Thomas Mikosch. 2006. "Quasi-maximum-likelihood estimation in conditionally heteroscedastic time series: A stochastic recurrence equations approach." *Annals of Statistics*, 34: 2449–2495.

Tong, Howell. 1983. *Threshold models in non-linear time series analysis.* Springer.

Tucker, Ledyard R. 1958. "An Inter-Battery Method of Factor Analysis." *Psychometrika*, 23: 111–136.

Wintenberger, Olivier. 2013. "Continuous invertibility and stable QML estimation of the EGARCH (1, 1) model." *Scandinavian Journal of Statistics*, 40: 846–867.

Zmijewski, Mark E. 1984. "Methodological issues related to the estimation of financial distress prediction models." *Journal of Accounting Research*, 22: 59–82.

Table 1: The Summary Statistics of Commercial Mortgages in the Top 10 MSA Areas

MSA Area	MSA Code	Mortgage Count	Percent of Count	Record Count	Percent of Records
Los Angeles-Long Beach-Anaheim, CA	31080	1,165	0.070	112,402	0.072
New York-Newark-Jersey City, NY-NJ-PA	35620	842	0.050	79,067	0.051
Dallas-Fort Worth-Arlington, TX	19100	631	0.038	58,247	0.037
Houston-The Woodlands-Sugar Land, TX	26420	530	0.032	48,395	0.031
Miami-Fort Lauderdale-West Palm Beach, FL	33100	478	0.029	46,986	0.030
Atlanta-Sandy Springs-Roswell, GA	12060	527	0.032	45,376	0.029
Phoenix-Mesa-Scottsdale, AZ	38060	473	0.028	44,764	0.029
Chicago-Naperville-Elgin, IL-IN-WI	16980	446	0.027	41,954	0.027
Philadelphia-Camden-Wilmington, PA-NJ-DE-MD	37980	354	0.021	34,600	0.022
Washington-Arlington-Alexandria, DC-VA-MD-WV	47900	348	0.021	34,013	0.022
Total	-	5,794	0.347	545,804	0.349

Notes: This table reports the summary statistics of commercial mortgages in the top 10 MSA areas.

Table 2: The Summary Statistics of Residential Mortgages in the Top 10 MSA Areas

MSA Area	MSA Code	Mortgage Count	Percent of Count	Record Count	Percent of Records
Los Angeles-Long Beach-Anaheim, CA	31080	1,582,224	0.065	46,588,138	0.043
New York-Newark-Jersey City, NY-NJ-PA	35620	1,131,311	0.047	52,777,972	0.048
Dallas-Fort Worth-Arlington, TX	19100	413,787	0.017	19,056,390	0.017
Houston-The Woodlands-Sugar Land, TX	26420	335,827	0.014	16,172,489	0.015
Miami-Fort Lauderdale-West Palm Beach, FL	33100	397,666	0.016	19,624,899	0.018
Atlanta-Sandy Springs-Roswell, GA	12060	446,344	0.018	21,513,426	0.020
Phoenix-Mesa-Scottsdale, AZ	38060	480,598	0.020	20,137,900	0.018
Chicago-Naperville-Elgin, IL-IN-WI	16980	884,902	0.037	35,582,091	0.032
Philadelphia-Camden-Wilmington, PA-NJ-DE-MD	37980	466,424	0.019	21,560,251	0.020
Washington-Arlington-Alexandria, DC-VA-MD-WV	47900	639,027	0.026	26,167,334	0.024
Total	-	6,778,110	0.280	279,180,890	0.255

Notes: This table reports the summary statistics of residential mortgages in the top 10 MSA areas.

Table 3: Components of Static and Dynamic Models for Commercial Mortgages

	Static I	Static II	Static III	Dynamic I	Dynamic II	Dynamic III
DSCR	X	X	X	X	X	X
Age 1	X	X	X	X	X	X
Age 2	X	X	X	X	X	X
Age 3	X	X	X	X	X	X
Age 4	X	X	X	X	X	X
Current HPI		X	X		X	X
HPI change		X	X		X	X
Current LTV		X	X		X	X
Anchor Flag			X			X
Shadow Anchor Flag			X			X
Single Tenant Flag			X			X
Acquisition Flag			X			X
Property Year			X			X
Property Age			X			X
Intercept	X	X	X			
Spatial Frailty				X	X	X

Notes: This table describes the model components of commercial mortgages in the empirical applications. X means the variable is used in the model.

Table 4: Components of Static and Dynamic Models for Residential Mortgages

	Static I	Static II	Static III	Dynamic I	Dynamic II	Dynamic III	Dynamic IV
DTI	X	X	X	X	X	X	X
Flag of first time home buyer	X	X	X	X	X	X	X
Subprime borrower	X	X	X	X	X	X	X
Loan purpose (non purchase)	X	X	X	X	X	X	X
Occupancy status (second home)	X	X	X	X	X	X	X
Occupancy status (investment)	X	X	X	X	X	X	X
Current LTV		X	X		X	X	X
HPI change		X	X		X	X	X
Current HPI		X	X		X	X	X
Property type (manufacturing housing)			X			X	X
Property type (Co-op)			X			X	X
Property type (PUD)			X			X	X
Property type (condo)			X			X	X
Channel (broker)			X			X	X
Channel (corresponent)			X			X	X
Combined LTV			X			X	X
Coborrower FICO score			X			X	X
Mortgage insurance (borrower paid)			X			X	X
Mortgage insurance (seller paid)			X			X	X
Mortgage insurance percentage			X			X	X
Count of borrowers			X			X	X
Count of units			X			X	X
Intercept	X	X	X				
Spatial Frailty				X	X	X	X
FICO Frailty							X

Notes: This table describes the model components of residential mortgages in the empirical applications. X means the variable is used in the model.

Table 5: Estimates of Static and Dynamic Residential Models in the LA Area

	Static I	Static II	Static III	Dynamic I	Dynamic II	Dynamic III	Dynamic IV
DTI	3.61	0.77	0.53	0.35	0.14	0.18	0.24
	(0.01)	(0.03)	(0.03)	(0.02)	(0.02)	(0.02)	(0.02)
Flag for first time home buyer	0.44	0.21	0.26	-0.05	-0.12	-0.11	-0.09
	(0.02)	(0.02)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)
Subprime borrower	0.11	0.11	0.09	0.08	0.09	0.10	0.41
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.06)
Loan purpose (non purchase)	0.31	0.11	0.13	-0.11	-0.05	-0.08	-0.05
	(0.01)	(0.01)	(0.01)	(7e-03)	(8e-03)	(8e-03)	(8e-03)
Occupancy status (second home)	-0.21	-0.03	-0.06	-0.32	-0.33	-0.28	-0.27
	(7e-03)	(8e-03)	(8e-03)	(0.02)	(0.02)	(0.02)	(0.02)
Occupancy status (investment)	-0.35	-0.31	-0.26	-0.35	-0.34	-0.31	-0.31
	(0.02)	(0.02)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)
Current LTV	-	0.32	0.31	-	0.57	0.42	0.10
		(0.01)	(0.01)		(0.02)	(0.03)	(0.04)
HPI change	-	-0.65	-0.04	-	-0.29	-0.76	-0.55
		(0.02)	(0.04)		(0.02)	(0.02)	(0.02)
Current HPI	-	-0.07	-0.36	-	-9e-03	-8e-03	-6e-03
		(0.01)	(0.02)		(2e-04)	(2e-04)	(2e-04)
Property type (manufacturing housing)	-	-	-0.01	-	-	-0.03	-0.09
			(7e-05)			(0.30)	(0.30)
Property type (Co-op)	-	-	-0.21	-	-	-0.90	-0.90
			(0.34)			(0.03)	(0.03)
Property type (PUD)	-	-	-0.85	-	-	0.04	0.05
			(0.03)			(7e-03)	(7e-03)
Property type (condo)	-	-	0.04	-	-	-0.11	-0.11
			(7e-03)			(7e-03)	(7e-03)
Channel (broker)	-	-	-0.09	-	-	0.19	0.20
			(7e-03)			(7e-03)	(7e-03)
Channel (corresponent)	-	-	0.18	-	-	0.13	0.13
			(7e-03)			(6e-03)	(6e-03)
Combined LTV	-	-	0.13	-	-	1.03	0.87
			(6e-03)			(0.03)	(0.03)
Coborrower FICO score	-	-	0.74	-	-	-0.09	-0.05
			(0.03)			(0.01)	(0.01)
Mortgage insurance (borrower paid)	-	-	-0.07	-	-	0.28	0.24
			(0.01)			(0.05)	(0.05)
Mortgage insurance (seller paid)	-	-	0.22	-	-	0.04	7e-03
			(0.05)			(0.04)	(0.04)
Mortgage insurance percentage	-	-	-0.02	-	-	0.13	0.21
			(0.04)			(0.18)	(0.18)
Count of borrowers	-	-	0.22	-	-	-0.16	-0.13
			(0.17)			(1e-02)	(1e-02)
Count of units	-	-	-0.14	-	-	-0.03	-0.02
			(9e-03)			(0.02)	(0.02)
θ	-	-	-	0.92	0.83	0.84	0.84
				(3e-03)	(5e-03)	(5e-03)	(4e-03)
α	-	-	-	0.18	0.18	0.18	0.18
				(5e-04)	(5e-04)	(5e-04)	(5e-04)
δ	-	-	-	-0.33	-0.26	-0.18	-0.34
				(0.01)	(0.01)	(0.01)	(0.01)
Intercept	-3.61	-0.77	-0.53	-	-	-	-
	(0.01)	(0.03)	(0.03)				
Log Likelihood	-5416021	-5195898	-5183015	-5103458	-5086902	-5072021	-1526772
AIC	10832148	10391948	10366379	10207052	10173987	10144421	3053925
BIC	10832873	10392984	10368762	10207985	10175230	10147011	3056310

Notes: This table reports the estimates of dynamic models for residential mortgages in the LA areas. The standard errors are reported in parenthesis.

Table 6: Estimates of Static and Dynamic Residential Models in the NYC Area

	Static I	Static II	Static III	Dynamic I	Dynamic II	Dynamic III	Dynamic IV
DTI	4.16	1.81	1.50	0.26	0.28	0.37	0.40
	(0.02)	(0.06)	(0.06)	(0.04)	(0.04)	(0.04)	(0.04)
Flag for first time home buyer	0.32	0.35	0.43	-0.06	-0.06	-0.06	-0.05
	(0.04)	(0.04)	(0.04)	(0.01)	(0.01)	(0.01)	(0.01)
Subprime borrower	0.12	0.08	0.07	0.33	0.34	0.33	0.40
	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)	(0.02)	(0.11)
Loan purpose (non purchase)	0.40	0.36	0.35	-0.15	-0.11	-0.07	-0.06
	(0.02)	(0.02)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)
Occupancy status (second home)	-0.22	-0.11	-0.07	-0.16	-0.16	-0.20	-0.20
	(0.01)	(0.01)	(0.01)	(0.05)	(0.05)	(0.05)	(0.05)
Occupancy status (investment)	-0.20	-0.16	-0.20	-0.15	-0.15	-0.12	-0.12
	(0.05)	(0.05)	(0.05)	(0.02)	(0.02)	(0.02)	(0.02)
Current LTV	-	0.18	0.16	-	0.30	0.22	0.14
		(0.02)	(0.02)		(0.03)	(0.05)	(0.05)
HPI change	-	-0.38	-0.05	-	-0.11	-0.12	-0.07
		(0.03)	(0.05)		(0.03)	(0.04)	(0.04)
Current HPI	-	-0.15	-0.08	-	-0.02	-0.02	-6e-03
		(0.02)	(0.03)		(7e-04)	(7e-04)	(7e-04)
Property type (manufacturing housing)	-	-	-0.01	-	-	0.01	0.61
			(2e-04)			(1.60)	(1.20)
Property type (Co-op)	-	-	0.93	-	-	-0.04	-0.03
			(1.01)			(0.01)	(0.01)
Property type (PUD)	-	-	-0.04	-	-	-0.31	-0.31
			(0.01)			(0.09)	(0.09)
Property type (condo)	-	-	-0.26	-	-	0.04	0.05
			(0.08)			(0.02)	(0.02)
Channel (broker)	-	-	0.03	-	-	0.18	0.18
			(0.01)			(0.01)	(0.01)
Channel (corresponent)	-	-	0.19	-	-	0.13	0.14
			(0.01)			(0.01)	(0.01)
Combined LTV	-	-	0.13	-	-	0.42	0.38
			(0.01)			(0.05)	(0.05)
Coborrower FICO score	-	-	0.33	-	-	-0.16	-0.15
			(0.05)			(0.02)	(0.02)
Mortgage insurance (borrower paid)	-	-	-0.24	-	-	0.19	0.17
			(0.02)			(0.12)	(0.12)
Mortgage insurance (seller paid)	-	-	0.15	-	-	0.11	0.10
			(0.12)			(0.06)	(0.06)
Mortgage insurance percentage	-	-	0.14	-	-	0.09	0.10
			(0.06)			(0.23)	(0.23)
Count of borrowers	-	-	-0.04	-	-	-0.17	-0.16
			(0.23)			(0.01)	(0.01)
Count of units	-	-	-0.21	-	-	-0.10	-0.10
			(0.01)			(0.01)	(0.01)
θ	-	-	-	0.93	0.91	0.91	0.90
				(8e-03)	(0.01)	(0.01)	(1e-02)
α	-	-	-	0.39	0.38	0.38	0.34
				(8e-03)	(8e-03)	(8e-03)	(7e-03)
δ	-	-	-	-0.29	-0.10	-0.05	-0.29
				(0.03)	(0.03)	(0.02)	(0.04)
Intercept	-4.16	-1.81	-1.50	-	-	-	-
	(0.02)	(0.06)	(0.06)				
Log Likelihood	-4075268	-4035582	-4029271	-4002358	-4000448	-3994682	-3141202
AIC	8150756	8071478	8059265	8004998	8001273	7990149	6283189
BIC	8152187	8073523	8063967	8006838	8003726	7995259	6288114

Notes: This table reports the estimates of dynamic models for residential mortgages in the NYC areas. The standard errors are reported in parenthesis.

Table 7: Estimates of Static and Dynamic Commercial Models in the LA Area

	Static I	Static II	Static III	Dynamic I	Dynamic II	Dynamic III
DSCR	-1.19	-1.08	-1.05	-1.19	-1.13	-1.06
	(0.07)	(0.08)	(0.09)	(23)	(0.39)	(0.09)
Age 1	0.03	0.03	0.04	0.02	0.03	0.04
	(3e-03)	(3e-03)	(3e-03)	(7.75)	(0.22)	(5e-03)
Age 2	-0.12	-0.08	-0.07	-0.11	-0.07	-0.07
	(0.01)	(0.01)	(0.02)	(10)	(0.02)	(0.01)
Age 3	0.78	0.81	0.82	0.75	0.77	0.79
	(0.06)	(0.06)	(0.06)	(28)	(0.24)	(0.06)
Current HPI	-	-0.01	-0.02	-	-0.02	-0.03
		(6e-03)	(7e-03)		(0.13)	(9e-03)
HPI change	-	-1.43	-1.44	-	-1.50	-1.50
		(0.19)	(0.20)		(0.33)	(0.21)
Current LTV	-	4.89	4.14	-	4.64	4.37
		(0.30)	(0.44)		(2.65)	(0.40)
Anchor Flag	-	-	0.19	-	-	0.22
			(0.18)			(0.24)
Shadow Anchor Flag	-	-	0.97	-	-	1.00
			(0.22)			(0.27)
Single Tenant Flag	-	-	1.41	-	-	1.42
			(0.16)			(0.19)
Acquisition Flag	-	-	1.08	-	-	1.06
			(0.22)			(0.21)
Property Year	-	-	9e-03	-	-	1e-03
			(0.04)			(1e-03)
Property Aage	-	-	0.02	-	-	0.02
			(0.04)			(3e-03)
θ	-	-	-	0.74	0.74	0.75
				(8.22)	(1.12)	(0.20)
α	-	-	-	0.95	0.94	0.92
				(27)	(0.98)	(0.22)
δ	-	-	-	-1.47	-1.56	-2.25
				(30)	(6.80)	(1.34)
Intercept	-6.09	-6.68	-24	-	-	-
	(0.19)	(0.45)	(83)			
Log Likelihood	-1677	-1542	-1465	-1594	-1513	-1440
AIC	3363	3099	2957	3202	3046	2911
BIC	3411	3176	3091	3269	3142	3064

Notes: This table reports the estimates of dynamic models for commercial mortgages in the LA areas. The standard errors for parameter estiamates are reported in parenthesis.

Table 8: Estimates of Static and Dynamic Commercial Models in the NYC Area

	Static I	Static II	Static III	Dynamic I	Dynamic II	Dynamic III
DSCR	-0.52	-0.41	-0.40	-0.54	-0.48	-0.50
	(0.11)	(0.14)	(0.14)	(0.28)	(0.87)	(0.58)
Age 1	6e-03	0.02	0.02	-1e-03	0.02	0.02
	(4e-03)	(4e-03)	(4e-03)	(0.05)	(0.17)	(0.05)
Age 2	-0.03	-6e-03	-0.03	-0.03	-9e-04	-0.03
	(0.01)	(0.01)	(0.02)	(0.07)	(0.49)	(0.10)
Age 3	0.48	0.55	0.76	0.46	0.54	0.73
	(0.04)	(0.04)	(0.06)	(0.31)	(1.81)	(0.50)
Current HPI	-	0.01	0.01	-	-0.01	-7e-03
		(7e-03)	(7e-03)		(0.05)	(0.10)
HPI change	-	-0.79	-0.73	-	-0.61	-0.52
		(0.24)	(0.24)		(1.08)	(3.66)
Current LTV	-	5.38	5.56	-	5.17	5.42
		(0.57)	(0.58)		(5.76)	(15)
Anchor Flag	-	-	-0.78	-	-	-0.77
			(0.23)			(0.71)
Shadow Anchor Flag	-	-	-0.03	-	-	-0.05
			(0.52)			(5.18)
Single Tenant Flag	-	-	-0.46	-	-	-0.42
			(0.26)			(0.26)
Acquisition Flag	-	-	-1.56	-	-	-1.58
			(0.72)			(6.38)
Property Year	-	-	3e-03	-	-	2e-03
			(6e-04)			(1e-03)
Property Aage	-	-	5e-03	-	-	5e-03
			(3e-03)			(3e-03)
θ	-	-	-	0.66	0.72	0.72
				(1.22)	(0.05)	(8.10)
α	-	-	-	1.28	1.04	1.04
				(0.70)	(0.52)	(0.95)
δ	-	-	-	-1.98	-2.75	-4.01
				(8.25)	(3.97)	(108)
Intercept	-6.10	-9.85	-15	-	-	-
	(0.24)	(0.71)	(1.42)			
Log Likelihood	-930	-873	-843	-898	-862	-831
AIC	1871	1763	1713	1810	1744	1695
BIC	1917	1836	1842	1875	1836	1842

Notes: This table reports the estimates of dynamic models for commercial mortgages in the NYC areas. The standard errors for parameter estiamates are reported in parenthesis.

Table 9: Casual Relationships among Spatial Frailty Factors for Commercial Mortgages

MSA	Atlanta	Chicago	Dallas	Houston	LA	Miami	NYC	Philadelphia	Phoenix	DC	Sign. Lags
Atlanta	-	0.42	<0.1	<0.01	<0.01	0.93	<0.05	0.12	<0.01	<0.01	5
Chicago	0.48	-	0.20	<0.01	<0.01	<0.01	<0.05	<0.1	0.50	0.13	4
Dallas	<0.1	<0.1	-	0.28	0.52	<0.05	0.16	0.52	<0.05	0.20	2
Houston	<0.01	0.61	<0.1	-	0.73	0.29	<0.01	0.94	<0.1	<0.01	3
LA	<0.05	0.80	<0.01	<0.1	-	0.10	0.59	0.29	<0.1	<0.01	3
Miami	<0.05	0.51	<0.05	<0.05	<0.01	-	<0.05	0.11	<0.01	<0.01	7
NYC	0.61	0.36	0.11	0.91	0.38	<0.01	-	0.11	0.17	0.10	1
Philadelphia	0.88	0.15	<0.01	<0.01	<0.01	<0.01	<0.01	-	0.39	0.22	5
Phoenix	<0.05	0.65	<0.01	0.13	<0.1	0.72	<0.05	0.65	-	<0.01	4
DC	<0.01	0.75	0.46	<0.1	0.46	<0.1	<0.01	<0.1	<0.1	-	2
Sign. Leads	5	0	4	4	4	4	7	0	3	5	-

Notes: The table reports the p-values of Granger causality tests between the frailty of commercial mortgages in main MSA areas. The length of lags included in causality tests is 6 months. The first column contains the dependent variables in the causality tests, and the first row contains the independent variable in the causality tests. The last row named 'Significant Leads' counts the number of p-values less than 0.05 in each column, and the last row named 'Significant Lags' counts the number of p-values less than 0.05 in each row.

Table 10: Granger Causality Tests between Prime Residential and Commercial Frailty

MSA	1 Month	3 Months	6 Months	12 Months	1 Month	3 Months	6 Months	12 Months
	Full sample				2000-2007 sample			
Atlanta	0.19	0.85	0.89	0.99	0.21	0.90	0.10	0.58
Chicago	0.01	0.02	0.08	0.82	0.01	0.03	0.01	0.13
Dallas	0.77	0.88	0.98	1.00	0.14	0.71	0.36	0.42
Houston	0.64	0.37	0.29	0.45	0.64	0.07	0.02	0.04
LA	0.26	0.75	0.84	0.56	0.77	0.21	0.03	1e-6
Miami	0.23	0.44	0.79	0.98	0.78	0.89	0.70	0.03
NYC	0.27	0.89	0.99	1.00	0.72	0.84	0.67	0.57
Philadelphia	0.76	0.95	0.83	0.29	0.63	0.78	0.05	0.13
Phoenix	0.90	0.94	0.98	1.00	1e-3	0.01	0.54	0.94
DC	0.60	0.86	0.98	1.00	0.02	0.03	0.01	1e-6

Notes: The table reports the p-values of Granger causality tests between commercial frailty and residential frailty in main MSA areas. The column title like '3 Month' indicates the number of lag variables used in the tests. Commercial frailty are the dependent variable and the frailty of prime borrowers are independent variables.

Table 11: Estimates of Joint Models for 10 MSA Areas

	Atlanta	Chicago	DC	Dallas	Houston	LA	Miami	NYC	Philadelphia	Phoenix
DSCR	-1.53	-0.91	-0.93	-1.26	-1.40	-1.06	-0.99	-0.47	-1.22	-1.28
	(0.13)	(0.09)	(0.28)	(0.84)	(0.13)	(0.09)	(0.12)	(0.13)	(0.14)	(0.09)
Age 1	0.02	0.03	0.04	0.02	0.04	0.04	0.02	0.02	0.02	6e-04
	(0.02)	(5e-03)	(0.01)	(0.02)	(0.03)	(4e-03)	(1e-02)	(0.10)	(6e-03)	(4e-03)
Age 2	-0.02	0.02	-0.01	0.02	-0.04	-0.07	2e-03	-0.03	5e-03	-0.04
	(0.01)	(9e-03)	(0.02)	(0.04)	(0.04)	(0.02)	(0.02)	(0.17)	(0.01)	(8e-03)
Age 3	0.63	0.21	0.60	0.14	0.48	0.79	0.45	0.79	0.23	0.20
	(0.07)	(0.03)	(0.07)	(0.08)	(0.05)	(0.06)	(0.04)	(0.08)	(0.06)	(0.03)
Current HPI	-0.02	-0.05	0.02	-0.04	-0.02	-0.03	-0.02	0.02	-0.06	-0.04
	(0.04)	(0.02)	(0.03)	(0.08)	(0.01)	(9e-03)	(0.02)	(0.18)	(0.02)	(9e-03)
HPI change	-0.28	0.02	-0.79	-0.31	-1.76	-1.49	-1.55	-0.52	-0.99	-1.11
	(0.34)	(0.29)	(0.43)	(0.85)	(0.33)	(0.20)	(0.29)	(0.24)	(0.49)	(0.20)
Current LTV	4.02	-0.19	4.59	1.68	5.52	4.40	3.62	6.01	0.45	1.09
	(0.38)	(0.41)	(1.19)	(1.50)	(0.62)	(0.40)	(1.04)	(0.65)	(0.72)	(0.17)
Anchor Flag	-0.38	0.19	-3.47	-0.62	-1.82	0.23	-0.41	-0.84	-0.60	0.56
	(0.12)	(0.15)	(0.88)	(0.22)	(0.28)	(0.19)	(0.22)	(0.22)	(0.24)	(0.10)
Shadow Anchor Flag	-0.06	-0.76	1.60	-0.61	-1.47	0.99	-1.29	-0.06	-200	0.17
	(0.17)	(0.30)	(0.42)	(1.38)	(0.29)	(0.20)	(0.81)	(0.47)	(52)	(0.13)
Single Tenant Flag	-16	-0.65	0.11	-3.38	-3.39	1.38	-18	-0.45	-1.00	-2.12
	(1.17)	(0.20)	(0.50)	(2.22)	(0.60)	(0.16)	(1.00)	(0.24)	(0.36)	(0.44)
Acquisition Flag	0.82	0.30	-15	1.13	-1.02	1.07	-14	-1.63	-361	-1.26
	(0.17)	(0.36)	(1.00)	(1.43)	(0.25)	(0.22)	(1.00)	(0.67)	(102)	(0.29)
Property Year	-4e-04	9e-03	2e-03	8e-03	4e-04	7e-03	-1e-04	3e-03	2e-03	7e-04
	(6e-04)	(8e-04)	(7e-03)	(2e-03)	(7e-04)	(5e-04)	(5e-04)	(5e-03)	(2e-03)	(6e-04)
Property Aage	-2e-03	3e-03	0.04	4e-03	-0.03	0.02	0.03	6e-03	0.02	-0.02
	(0.02)	(3e-03)	(9e-03)	(0.01)	(0.02)	(2e-03)	(5e-03)	(0.12)	(3e-03)	(5e-03)
θ_{cr}	0.63	0.40	-1e-02	-1e-02	1.01	0.09	-1e-02	-1e-02	0.12	0.72
	(0.26)	(0.27)	(0.81)	(2.52)	(0.30)	(0.26)	(0.59)	(0.46)	(0.34)	(0.36)
θ_{cc}	0.87	0.81	0.69	0.87	0.87	0.76	0.81	0.73	0.55	0.95
	(0.12)	(0.10)	(0.18)	(0.24)	(0.10)	(0.08)	(0.47)	(0.11)	(0.14)	(0.01)
α	0.43	0.57	1.17	0.44	0.45	0.88	0.70	1.00	1.14	0.13
	(0.05)	(0.08)	(0.30)	(0.85)	(0.15)	(0.19)	(0.38)	(0.27)	(0.23)	(0.02)
δ	-0.60	-3.77	-4.71	-2.49	-0.50	-5.03	-1.08	-4.88	-1.66	0.11
	(0.52)	(1.94)	(4.35)	(4.72)	(0.36)	(1.69)	(2.79)	(1.84)	(2.18)	(0.11)

Notes: This table reports the estimates of joint models for commercial mortgages in the 10 MSA areas. The standard errors are reported in parenthesis. Note that we omit the estimates for residential mortgages because they are similar to the ones in Dynamic III.

Figure 1: Comparison of Fitted PD by Static and Dynamic Models
Residential Mortgages in New York City, DC, Dallas and Phoenix

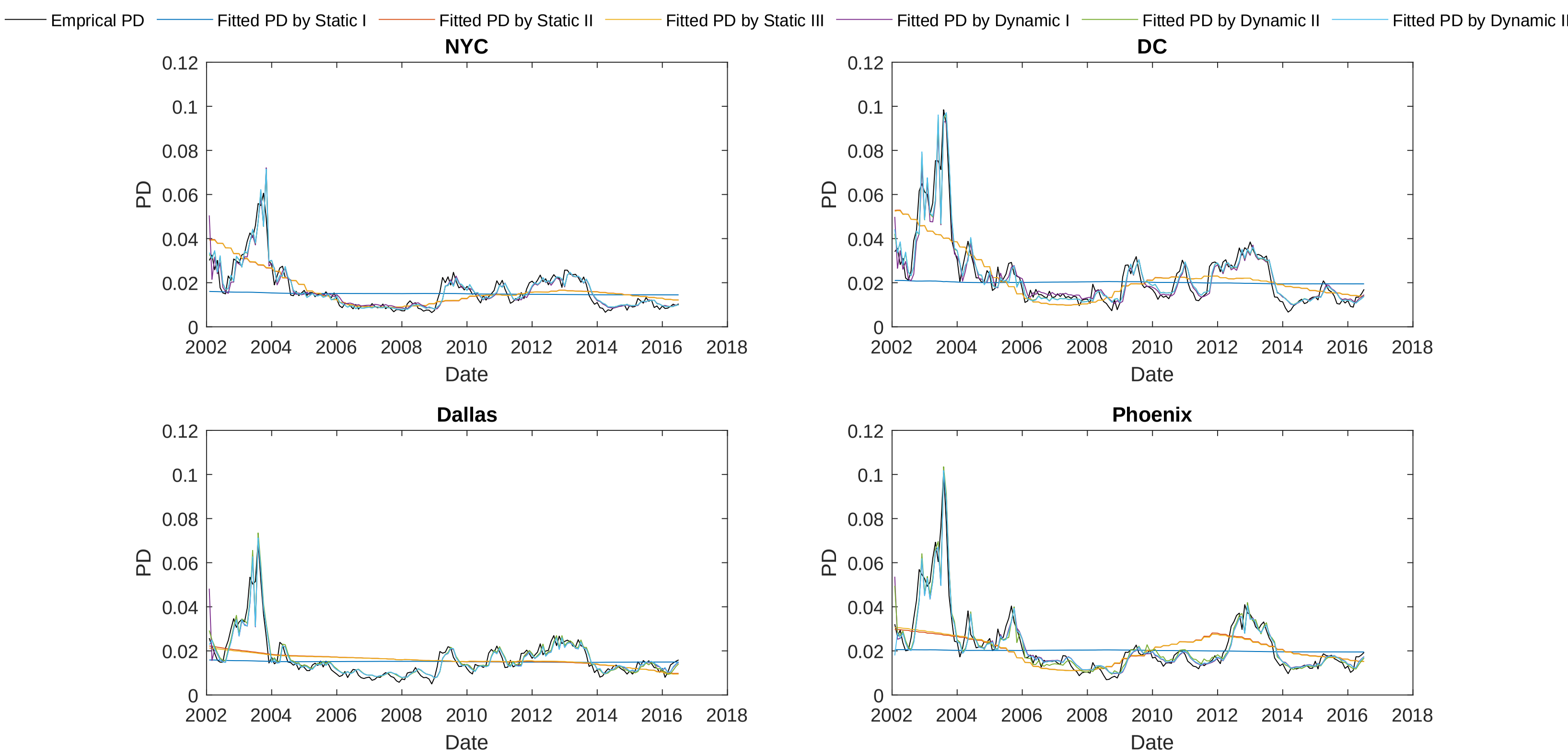
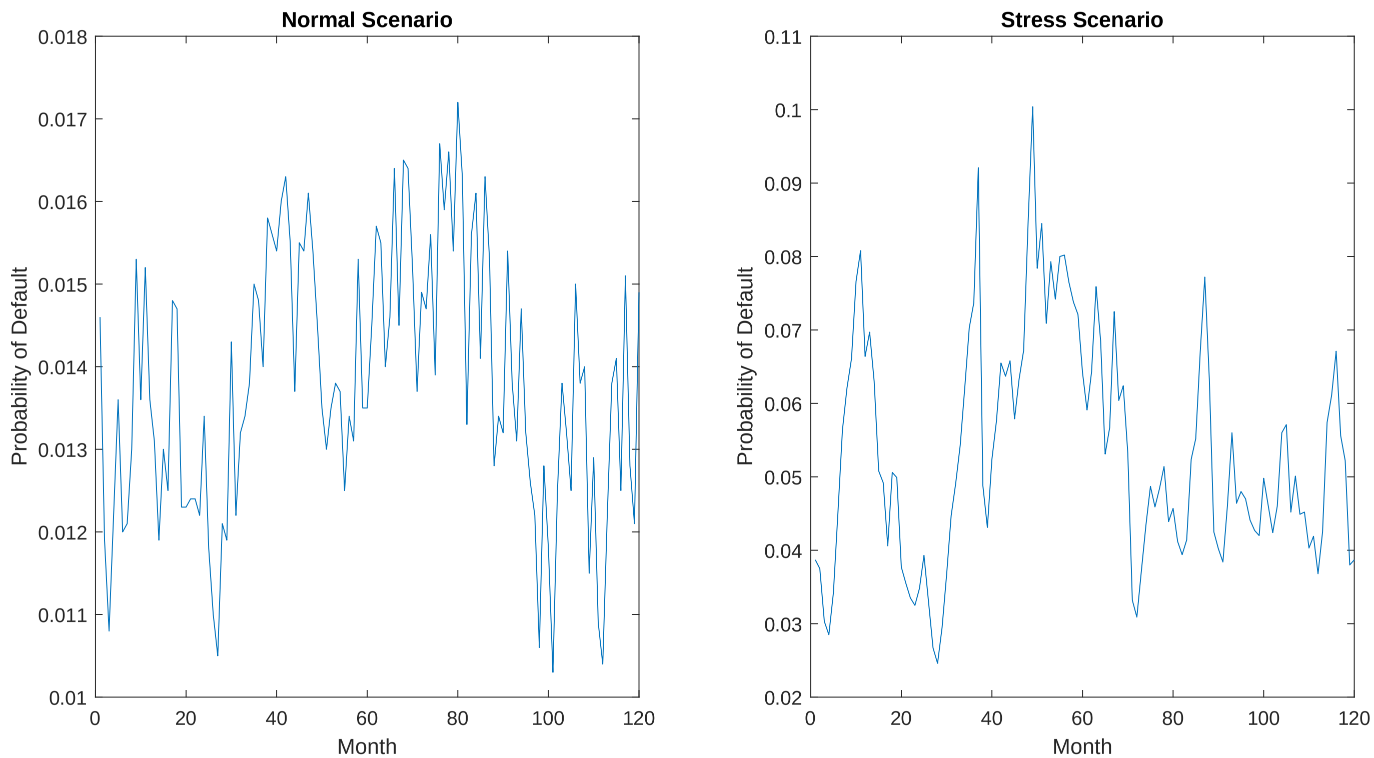


Figure 2: Stress testing simulations



APPENDIX

A.1 Asymptotic properties of the QMLE

To obtain the asymptotic properties of the QMLE estimator we need stationarity, ergodicity, and some mixing conditions on stochastic processes generated by the model. For univariate GAS models Blasques, Koopman and Lucas (2014) provide a rigorous analysis. Consider the binary choice representation of the model

$$\begin{aligned} y_{i,t} &= \mathbb{1}\left\{x_{i,t}^\top \beta + c_i^\top f_t \geq u_{i,t}\right\} \\ \pi_{i,t}(\phi, f_t) &= \frac{1}{1+e^{-(x_{i,t}^\top \beta + c_i^\top f_t)}} \\ f_t &= \delta + \theta f_{t-1} + \alpha s_{t-1}(\phi, f_{t-1}) \\ s_{t-1}(\phi, f_{t-1}) &= \left(\frac{1}{N_j^k}\sum_{i=1}^n (y_{i,t-1} - \pi_{i,t-1}(\phi, f_{t-1}))\mathbb{1}\left\{c_i^k = \iota_j^k\right\}\right)_{\substack{j=1,\dots,d_k \\ k=1,\dots,d}}, \end{aligned}$$

where $u_{i,t}$ are i.i.d. random variables, $x_{i,t}$ are covariates, ι_j^k is a d_k-dimensional vector for which the jth entry is one and the rest are zero.
For simplicity suppose that $n_t = n$ for all t. Let $\phi = (\alpha, \beta, \delta, \theta)$ be a vector of parameters in some parameter space $\Phi \subset \mathbf{R}^{2d^2+d+L}$. We are interested in ϕ_0 solving

$$\phi_0 = \arg\max_{\phi \in \Phi} \mathbb{E}[l_t(\phi)],$$

where

$$\begin{aligned} l_t(\phi) &\triangleq \frac{1}{n}\sum_{i=1}^n \left\{y_{i,t}\left(x_{i,t}^\top \beta + c_i^\top f_t\right) - \ln\left(1 + e^{x_{i,t}^\top \beta + c_i^\top f_t}\right)\right\} \\ &\triangleq \frac{1}{n}\sum_{i=1}^n l_{i,t}(\phi). \end{aligned}$$

The empirical counterpart to this problem is to maximize

$$\mathcal{L}_T(\phi) = \frac{1}{T}\sum_{t=1}^T l_t(\phi). \tag{A.1}$$

Note that the likelihood in (A.1) is not accessible in practice, because the factor process $(f_t)_{t\in\mathbf{Z}_+}$ is not observed. It can be computed recursively given its initial value is drawn from some distribution, e.g. the degenerate distribution with point mass at some value $\tilde{f}_0 \in \mathbf{R}^d$. Let $(\tilde{f}_t)_{\mathbf{Z}_+}$ be the factor process initiated from $\tilde{f}_0$ and $(f_t)_{t\in\mathbf{Z}_+}$ be the factor process initiated from its stationary distribution. Similarly, $\tilde{l}_{it}, \tilde{\mathcal{L}}_T$ are computed based on $\tilde{f}_t$ initiated from $\tilde{f}_0$.
In the Lemma A.1.1 below we show that the choice of the initial value does not matter asymptotically. To that end, we need to introduce several Assumptions.

Assumption A.1.1. *(i) for any $t \in \mathbf{Z}_+$, $(x_t, u_t) = (x_{i,t}, u_{i,t})_{i=1}^n$ is a collection of i.i.d. random vectors; (ii) $(x_{i,t})_{t\in\mathbf{Z}_+}$ is a stationary geometrically ergodic Markov process with respect to the filtration of the factor process; (iii) $(u_t)_{t\in\mathbf{Z}_+}$ are i.i.d. and independent from $(x_t)_{t\in\mathbf{Z}_+}$ (iv) $(c_i)_{i=1}^n$ are predetermined.*

Note that the above Assumption requires i.i.d. sampling in the cross-sectional dimension and allows for dependent data in the time series dimension.

Assumption A.1.2. *(i)* $(x_{i,0}^\top\beta_0, \ldots, x_{i,t}^\top\beta_0)$ *have strictly positive Lebesgue density on* $\mathbf{R}^{t+1}$ *for any* $t \in \mathbf{Z}_+$*; (ii)* $u_{i,t}$ *have continuous strictly positive Lebesgue density on* $\mathbf{R}$*; (iii)* $\mathbb{E}\|x_{i,t}\|^{2+\epsilon} < \infty$ *for some* $\epsilon > 0$.

The above Assumption serves a few purposes, including verification of the geometric ergodicity of the factor process. (i) and (ii) are relatively mild conditions. For example, (i) does not rule-out the possibility of having discrete regressors in $x_{i,t}$. However, at least one of covariates should have a Lebesgue probability density function. (ii) is obviously satisfied if $u_{i,t}$ has logistic distribution. (iii) requires existence of moments of order slightly larger than 2 and is needed for the mixing CLT.
The last Assumption provides sufficient conditions for invertibility of the model. It is beyond the scope of the current paper to establish both necessary and sufficient conditions, which might be challenging given the complex non-linear dynamics exhibited by our model.

Assumption A.1.3. *(i) For some constants* c *and* b

$$\Phi \subset \left\{\phi = (\alpha, \beta, \delta, \theta) : \ \rho(\theta) + \frac{d}{4}\rho(\alpha) \leq c < 1, \ \rho(\theta) \geq b > 0\right\},$$

where for a matrix A*,* $\rho(A)$ *denotes its spectral radius; (ii)* Φ *is compact and* ϕ_0 *is an interior point of* Φ*; (iii)* $\lambda_{\min}(\theta^\top\theta) > 0$*,* $\lambda_{\min}(\alpha^\top\alpha) > 0$*, and* $\lambda_{\min}(J(\phi)) < 0$ *for all* $\phi \in \Phi$*, where* $\lambda_{\min}(A)$ *denotes the smallest eigenvalue of* A.

Note that the score[13] and the Hessian for the observation (i, t) have the following expressions

$$\begin{aligned}
\nabla l_{i,t}(\phi) &= (y_{i,t} - \pi_{i,t}(\phi, f_t))\xi_{i,t}(\phi) \\
\nabla^2 l_{i,t}(\phi) &= -\pi'\left(x_{i,t}^\top\beta + c_i^\top f_t\right)\xi_{i,t}(\phi)\xi_{i,t}^\top(\phi) + (y_{i,t} - \pi_{i,t}(\phi, f_t))\frac{\partial \xi_{i,t}(\phi)}{\partial \phi^\top},
\end{aligned}$$

where

$$\xi_{i,t}(\phi) = \nabla(x_{i,t}^\top\beta + c_i^\top f_t(\phi)) \tag{A.2}$$

and π' is the Lebesgue density of the Logistic distribution.

Lemma A.1.1. *Suppose that Assumption A.1.3 is satisfied and that* $\tilde{f}_0$ *is drawn from the distribution with a compact support that does not depend on the parameter* ϕ*. Then*

$$\sup_{\phi\in\Phi}\left|\tilde{\mathcal{L}}_T(\phi) - \mathcal{L}_T(\phi)\right| = O_{\text{a.s.}}(T^{-1}).$$

Proof. First we bound the difference between the sample log-likelihood with the initial value $\tilde{f}_0$ and the initial value drawn from the stationary distribution as follows

$$\begin{aligned}
&\left|\tilde{\mathcal{L}}_T(\phi) - \mathcal{L}_T(\phi)\right| \\
&\leq \frac{1}{Tn}\sum_{t=1}^{T}\sum_{i=1}^{n}\left|y_{i,t}\ln\left(\frac{1 + e^{-(x_{i,t}^\top\beta + c_i^\top f_t)}}{1 + e^{-(x_{i,t}^\top\beta + c_i^\top \tilde{f}_t)}}\right) + (1 - y_{i,t})\ln\left(\frac{1 + e^{x_{i,t}^\top\beta + c_i^\top f_t}}{1 + e^{x_{i,t}^\top\beta + c_i^\top \tilde{f}_t}}\right)\right|.
\end{aligned}$$

[13] Derivatives with respect to matrices α and θ are understood as derivatives with respect to vectorizations $\mathrm{vec}(\alpha)$ and $\mathrm{vec}(\theta)$.

Using the inequality $\left|\ln\left(\frac{1+e^x}{1+e^y}\right)\right| \leq |x-y|$ for all $x, y \in \mathbf{R}$ and $y_{i,t} \in \{0,1\}$

$$\begin{aligned}\left|\tilde{\mathcal{L}}_T(\phi) - \mathcal{L}_T(\phi)\right| &\leq \frac{1}{Tn}\sum_{t=1}^{T}\sum_{i=1}^{n}\left|c_i^\top(\tilde{f}_t(\phi) - f_t(\phi))\right| \\ &\leq d^{1/2}\frac{1}{T}\sum_{t=1}^{T}\left\|\tilde{f}_t(\phi) - f_t(\phi)\right\|,\end{aligned}$$

where the second line follows by the Cauchy-Schwartz inequality and $\|c_i\| \leq d^{1/2}$. Iterating the factor process backwards

$$f_t = (I-\theta)^{-1}(I-\theta^t)\delta + \theta^t f_0 + \sum_{j=1}^{t}\theta^{j-1}\alpha s_{t-j}(\phi, f_{t-j}). \tag{A.3}$$

Then by the triangle inequality

$$\left\|\tilde{f}_t - f_t\right\| \leq \rho(\theta)^t\left\|\tilde{f}_0 - f_0\right\| + \rho(\alpha)\sum_{j=1}^{t}\rho(\theta)^{j-1}\left\|s_{t-j}(\phi, \tilde{f}_{t-j}) - s_{t-j}(\phi, f_{t-j})\right\|$$

and by the Cauchy-Schwartz inequality

$$\begin{aligned}\left\|s_t(\phi, \tilde{f}_t) - s_t(\phi, f_t)\right\| &\leq d^{1/2}\max_{j,k}\left|\frac{1}{N_j^k}\sum_{i=1}^{n}\left(\pi_{i,t}(\phi, f_t) - \pi_{i,t}(\phi, \tilde{f}_t)\right)\mathbb{1}\{c_i^k = \iota_j^k\}\right| \\ &\leq d^{1/2}\max_{1\leq i\leq N}\left|\pi_{i,t}(\phi, \tilde{f}_t) - \pi_{i,t}(\phi, f_t)\right| \\ &\leq \frac{d^{1/2}}{4}\max_{1\leq i\leq N}\left|c_i^\top(\tilde{f}_t - f_t)\right| \\ &\leq \frac{d}{4}\left\|\tilde{f}_t - f_t\right\|,\end{aligned}$$

where we use the mean-value theorem, the fact that the Logistic density is uniformly bounded by 1/4, and the Cauchy-Schwartz inequality. Therefore,

$$\left\|\tilde{f}_t - f_t\right\| \leq \rho(\theta)^t\left\|\tilde{f}_0 - f_0\right\| + \frac{d}{4}\rho(\alpha)\sum_{j=1}^{t}\rho(\theta)^{j-1}\left\|\tilde{f}_{t-j}(\phi) - f_{t-j}(\phi)\right\|.$$

Iterating this inequality recursively and grouping all terms

$$\left\|\tilde{f}_t - f_t\right\| \leq \left\|\tilde{f}_0 - f_0\right\|\left(\rho(\theta) + \frac{d}{4}\rho(\alpha)\right)^t, \tag{A.4}$$

whence under Assumption A.1.3 (i) and the triangle inequality

$$\left|\tilde{\mathcal{L}}_T(\phi) - \mathcal{L}_T(\phi)\right| \leq d^{1/2}\left\|\tilde{f}_0 - f_0\right\|\frac{1}{T}\sum_{t=1}^{T}c^t \leq \frac{1}{T}\frac{d^{1/2}}{1-c}\left(\|\tilde{f}_0\| + \|f_0\|\right).$$

Finally, by the compact support of $\tilde{f}_0$, Proposition A.2.1 ,and Assumption A.1.3 (i)

$$\sup_{\phi\in\Phi}\left|\tilde{\mathcal{L}}_T(\phi) - \mathcal{L}_T(\phi)\right| = O_{\text{a.s.}}\left(\frac{1}{T}\right).$$

□

Lemma A.1.2. *Suppose that Assumption A.1.3 is satisfied, $\mathbb{E}\|x_{i,t}\| < \infty$, and that $\tilde{f}_0$ is drawn from the distribution with compact support that does not depend on the parameter ϕ. Then*[14]

$$\sup_{\phi \in \Phi} \left\| \nabla \tilde{\mathcal{L}}_T(\phi) - \nabla \mathcal{L}_T(\phi) \right\| = \mathrm{O_p}(T^{-1}).$$

Proof. By the triangle inequality

$$\begin{aligned} \left\| \nabla \tilde{\mathcal{L}}_T(\phi) - \nabla \mathcal{L}_T(\phi) \right\| &= \left\| \frac{1}{Tn} \sum_{i=1}^{n} \sum_{t=1}^{T} \nabla \tilde{l}_{i,t}(\phi) - \nabla l_{i,t}(\phi) \right\| \\ &\leq \frac{1}{Tn} \sum_{i=1}^{n} \sum_{t=1}^{T} \left\| \nabla \tilde{l}_{i,t}(\phi) - \nabla l_{i,t}(\phi) \right\|. \end{aligned}$$

Next by the triangle inequality

$$\begin{aligned} \left\| \nabla \tilde{l}_{i,t}(\phi) - \nabla l_{i,t}(\phi) \right\| &= \left\| (y_{i,t} - \tilde{\pi}_{i,t}(\phi)) \left(\tilde{\xi}_{i,t}(\phi) - \xi_{i,t}(\phi) \right) + (\pi_{i,t}(\phi) - \tilde{\pi}_{i,t}(\phi)) \xi_{i,t}(\phi) \right\| \\ &\leq \left\| \tilde{\xi}_{i,t}(\phi) - \xi_{i,t}(\phi) \right\| + |\pi_{i,t}(\phi) - \tilde{\pi}_{i,t}(\phi)| \, \|\xi_{i,t}(\phi)\|. \end{aligned}$$

Since the density of the Logistic distribution is uniformly bounded by $1/4$, by the mean-value theorem and the Cauchy-Schwartz inequality

$$\begin{aligned} |\pi_{i,t}(\phi) - \tilde{\pi}_{i,t}(\phi)| &\leq \frac{1}{4} \left| c_i^\top (f_t - \tilde{f}_t) \right| \\ &\leq \frac{d^{1/2}}{4} \left\| \tilde{f}_t - f_t \right\|. \end{aligned}$$

Using the definition of $\xi_{i,t}(\phi)$ in (A.2)

$$\begin{aligned} \|\xi_{i,t}(\phi)\| &\leq \|\nabla x_{i,t}^\top \beta\| + \|\nabla c_i^\top f_t\| \\ &= \|x_{i,t}\| + \left\| (J_\phi f_t(\phi))^\top c_i \right\| \\ &\leq \|x_{i,t}\| + d^{1/2} \left\| J_\phi f_t(\phi) \right\|, \end{aligned}$$

where J_ϕ denotes the Jacobian with respect to the parameter ϕ. Note that for partitioned matrix $(A \; B)$, the spectral norm obeys $\|(A \; B)\| \leq \|A\| + \|B\|$, so in order to find bound the norm of the Jacobian $J_\phi f_t(\phi)$, it is sufficient to consider individual norm bounds on Jacobians with respect to subvectors

$$\begin{aligned} J_\alpha f_t &= (\theta + \alpha J_f s_{t-1}(f_{t-1})) J_\alpha f_{t-1} + S_{t-1} \\ J_\beta f_t &= \theta J_\beta f_{t-1} + \alpha J_\beta s_{t-1}(f_{t-1}) \\ J_\delta f_t &= (\theta + \alpha J_f s_{t-1}(f_{t-1})) J_\delta f_{t-1} + I_d \\ J_\theta f_t &= (\theta + \alpha J_f s_{t-1}(f_{t-1})) J_\theta f_{t-1} + F_{t-1}, \end{aligned}$$

where $S_t = (\mathrm{diag}(s_{t,1}), \ldots, \mathrm{diag}(s_{t,d}))$ and $F_t = (\mathrm{diag}(f_{t,1}), \ldots, \mathrm{diag}(f_{t,d}))$. The term $J_f s_t$ appears in three Jacobians, so we bound its norm first. Note that for a $d \times d$ matrix A, we

[14] Note that by continuity, the supremum is taken over the countable set of rationals, whence it is measurable. We will use this fact repeatedly in proofs.

have $\|A\| \le d \max_{i,j} |A_{i,j}|$, whence

$$\begin{aligned}
\|J_f s_t(f_t)\| &= \left\| \left(-\frac{1}{N_j^k} \sum_{i=1}^{n} \mathbb{1}\{c_i^k = \iota_j^k\} \pi'_{i,t}(f) c_{i,r} \right)_{\substack{j,k \\ r}} \right\| \\
&\le d \max_{j,k,r} \left| \frac{1}{N_j^k} \sum_{i=1}^{n} \mathbb{1}\{c_i^k = \iota_j^k\} \pi'_{i,t}(f) c_{i,r} \right| \\
&\le \frac{d}{4},
\end{aligned}$$

where we use the fact that the density of the Logistic distribution is bounded by 1/4. Then in the stationary regime

$$\begin{aligned}
\|J_\alpha f_t\| &\le (\rho(\theta) + \rho(\alpha) d/4) \|J_\alpha f_{t-1}\| + d\|s_{t-1}\|_\infty \\
&\le c\|J_\alpha f_{t-1}\| + d \\
&\le \frac{d}{1-c} < \infty.
\end{aligned}$$

Similarly, in the stationary regime we obtain

$$\|J_\delta f_t\| \le \frac{d}{1-c}$$

and

$$\begin{aligned}
\|J_\theta f_t\| &\le \frac{d}{1-c} \|f_t\|_\infty \\
&= \frac{d}{1-c} \frac{\|\delta\| + d^{1/2}\rho(\alpha)}{1-\rho(\theta)},
\end{aligned}$$

where the last line follows by Proposition A.2.1.
Lastly

$$\|J_\beta f_t\| \le \rho(\theta)\|J_\beta f_{t-1}\| + \rho(\alpha) \|J_\beta s_{t-1}(f_{t-1}(\phi))\|,$$

where by the triangle inequality, Jensen's inequalities, and $\|Jc\| \le \|J\|\|c\|$

$$\begin{aligned}
\|J_\beta s_t(f_t(\phi))\| &= \left\| \left(\frac{1}{N_j^k} \sum_{i=1}^{n} \mathbb{1}\{c_i^k = \iota_j^k\} \pi'_{i,t}(\phi)(x_{i,t}^\top + c_i^\top J_\beta f_t) \right)_{j,k} \right\| \\
&\le \left\| \left(\frac{1}{N_j^k} \sum_{i=1}^{n} \mathbb{1}\{c_i^k = \iota_j^k\} \pi'_{i,t}(\phi) x_{i,t}^\top \right)_{j,k} \right\| \\
&\quad + \left\| \left(\frac{1}{N_j^k} \sum_{i=1}^{n} \mathbb{1}\{c_i^k = \iota_j^k\} \pi'_{i,t}(\phi) c_i^\top J_\beta f_t \right)_{j,k} \right\| \\
&\le \frac{1}{4} \sqrt{\sum_{j,k} \frac{1}{N_j^k} \sum_{i=1}^{n} \mathbb{1}\{c_i^k = \iota_j^k\} \|x_{i,t}\|^2} \\
&\quad + \frac{1}{4} \sqrt{\sum_{j,k} \frac{1}{N_j^k} \sum_{i=1}^{n} \mathbb{1}\{c_i^k = \iota_j^k\} \|(J_\beta f_t)^\top c_i\|^2} \\
&\le \frac{d}{4} \|J_\beta f_t\| + \frac{1}{4} \sqrt{\sum_{j,k} \frac{1}{N_j^k} \sum_{i=1}^{n} \mathbb{1}\{c_i^k = \iota_j^k\} \|x_{i,t}\|^2}.
\end{aligned}$$

Combining above estimates

$$\|J_\beta f_t(\phi)\| \le c\|J_\beta f_{t-1}\| + \frac{1}{4}\sqrt{\sum_{j,k}\frac{1}{N_j^k}\sum_{i=1}^n \mathbb{1}\{c_i^k = \iota_j^k\}\|x_{i,t}\|^2}.$$

This implies that

$$\sup_{\phi\in\Phi}\|J_\beta f_t(\phi)\| \le \frac{1}{4}\sum_{j=0}^\infty c^j\sqrt{\sum_{j,k}\frac{1}{N_j^k}\sum_{i=1}^n \mathbb{1}\{c_i^k = \iota_j^k\}\|x_{i,t-j}\|^2},$$

whence by stationarity and Jensen's inequality

$$\mathbb{E}\left[\sup_{\phi\in\Phi}\|J_\beta f_t(\phi)\|\right] \le \frac{\sqrt{\mathbb{E}\|x_{i,t}\|^2}}{1-c}, \tag{A.5}$$

implying

$$\sup_{\phi\in\Phi}\|J_\beta f_t(\phi)\| = \mathrm{O_p}(1).$$

Combining all above computations with inequality in (A.4), we obtain uniformly over $\phi \in \Phi$

$$\begin{aligned}\left\|\nabla\tilde{\mathcal{L}}_T(\phi) - \nabla\mathcal{L}_T(\phi)\right\| &\le \frac{d^{1/2}}{4}\left\|\tilde{f}_0 - f_0\right\|\left(\frac{1}{Tn}\sum_{i,t}c^t\|x_{i,t}\| + \mathrm{O_p}(T^{-1})\right) + \frac{1}{Tn}\sum_{i,t}\left\|\tilde{\xi}_{i,t}(\phi) - \xi_{i,t}(\phi)\right\| \\ &= \frac{1}{Tn}\sum_{i,t}\left\|\tilde{\xi}_{i,t}(\phi) - \xi_{i,t}(\phi)\right\| + \mathrm{O_p}(T^{-1}),\end{aligned}$$

where the second line follows by Markov's inequality and Assumption A.1.2 (iii) and the $\mathrm{O_p}$ term does not depend on ϕ. It remains to show that the first terms is also bounded in probability with rate T^{-1}. To that end, note that

$$\begin{aligned}\left\|\tilde{\xi}_{i,t}(\phi) - \xi_{i,t}(\phi)\right\| &= \left\|(J_\phi\tilde{f}_t(\phi) - J_\phi f_t(\phi))^\top c_i\right\| \\ &\le d^{1/2}\left\|J_\phi\tilde{f}_t(\phi) - J_\phi f_t(\phi)\right\|,\end{aligned}$$

where

$$\begin{aligned}\left\|J_\phi\tilde{f}_t(\phi) - J_\phi f_t(\phi)\right\| &\le \rho(\theta)\left\|J_\phi\tilde{f}_{t-1}(\phi) - J_\phi f_{t-1}(\phi)\right\| \\ &\quad + \rho(\alpha)\left\|J_\phi s_{t-1}(\tilde{f}_{t-1}) - J_\phi s_{t-1}(f_{t-1})\right\|\end{aligned}$$

and

$$\begin{aligned}\left\|J_\phi s_{t-1}(\tilde{f}_{t-1}) - J_\phi s_{t-1}(f_{t-1})\right\| &= \left\|\left(\frac{1}{N_j^k}\sum_{i=1}^n \mathbb{1}\{c_i^k = \iota_j^k\}\frac{\partial}{\partial\phi^\top}\left\{\pi_{i,t-1}(\tilde{f}_{t-1}) - \pi_{i,t-1}(f_{t-1})\right\}\right)_{j,k}\right\| \\ &\le d^{1/2}\max_{1\le i\le N}\left\|\nabla_\phi\left(\pi_{i,t-1}(\tilde{f}_{t-1}) - \pi_{i,t-1}(f_{t-1})\right)\right\|.\end{aligned}$$

The last norm is bounded as follows

$$\begin{aligned}\left\|\nabla_\phi\left(\pi_{i,t-1}(\tilde{f}_{t-1}) - \pi_{i,t-1}(f_{t-1})\right)\right\| &= \left\|\pi'_{i,t-1}(\tilde{f}_{t-1})\tilde{\xi}_{i,t-1} - \pi'_{i,t-1}(f_{t-1})\xi_{i,t-1}\right\| \\ &\le \frac{1}{4}\left\|\tilde{\xi}_{i,t-1} - \xi_{i,t-1}\right\| + \left|\pi'_{i,t-1}(\tilde{f}_{t-1}) - \pi'_{i,t-1}(f_{t-1})\right|\|\xi_{i,t-1}\| \\ &\le \frac{d^{1/2}}{4}\left\|J_\phi\tilde{f}_{t-1}(\phi) - J_\phi f_{t-1}(\phi)\right\| \\ &\quad + 0.01\left\|\tilde{f}_{t-1} - f_{t-1}\right\|\{\mathrm{O_p}(1) + \|x_{i,t-1}\|\}\end{aligned}$$

Therefore by inequality in the (A.4)

$$\begin{aligned}\left\|J_\phi \tilde{f}_t(\phi) - J_\phi f_t(\phi)\right\| &\leq c\left\|J_\phi \tilde{f}_{t-1}(\phi) - J_\phi f_{t-1}(\phi)\right\| + \mathrm{O_p}(c^t)\\ &= c\left\|J_\phi \tilde{f}_{t-1}(\phi) - J_\phi f_{t-1}(\phi)\right\| + \mathrm{O_p}(c^{t/2}),\end{aligned}$$

where the $\mathrm{O_p}$ term does not depend on ϕ by Proposition A.2.1 and restrictions of Φ. This shows that

$$\left\|J_\phi \tilde{f}_t(\phi) - J_\phi f_t(\phi)\right\| = \mathrm{O_p}\left(\sum_{j=0}^{\infty} c^j c^{(t-j)/2}\right) = \mathrm{O_p}(c^{t/2})$$

uniformly over $\phi \in \Phi$. This shows that

$$\sup_{\phi\in\Phi} \frac{1}{Tn} \sum_{i,t} \left\|\tilde{\xi}_{i,t}(\phi) - \xi_{i,t}(\phi)\right\| = \mathrm{O_p}\left(T^{-1}\sum_{t=1}^{T} c^{t/2}\right)$$

and completes the proof. □

By Lemma A.1.1 we can define the QMLE $\hat{\phi}_T$ as a solution to the approximate maximization problem

$$\begin{aligned}\hat{\phi}_T &= \arg\max_{\phi\in\Phi} \tilde{\mathcal{L}}_T(\phi)\\ &= \arg\max_{\phi\in\Phi} \mathcal{L}_T(\phi) + \mathrm{o_{a.s.}}(1)\end{aligned}$$

If the logistic specification is correct, then $\hat{\phi}_T$ coincides with the MLE.

Theorem A.2. *Suppose that Assumptions of Lemma A.1.1 are satisfied. Then under Assumptions A.1.1, A.1.2, and A.1.3 the QMLE estimator is strongly consistent, as $T \to \infty$*

$$\hat{\phi}_T \xrightarrow{\text{a.s.}} \phi_0.$$

Proof. First, we note that the continuity of the log-likelihood and the compactness of Φ ensure that $\hat{\phi}_T$ is measurable.
Let $\mathbb{E}_T$ be the empirical expectation operator, so that $\mathbb{E}_T l_t = \frac{1}{T}\sum_{t=1}^{T} l_t(y_t, x_t, f_t)$ and let $\mathbb{E}$ be the population expectation operator with respect to (y_t, x_t, f_t) only, so that $\mathbb{E}l_t(\hat{\phi}_T) = \int l(y, x, f; \hat{\phi}_T)P(\mathrm{d}y, \mathrm{d}x, \mathrm{d}f)$. Then

$$\begin{aligned}0 &\leq \mathbb{E}l_t(\phi_0) - \mathbb{E}l_t(\hat{\phi}_T)\\ &= \left(\mathbb{E}_T l_t(\hat{\phi}_T) - \mathbb{E}l_t(\hat{\phi}_T)\right) - (\mathbb{E}_T l_t(\phi_0) - \mathbb{E}l_t(\phi_0)) + \left(\mathcal{L}_T(\phi_0) - \mathcal{L}_T(\hat{\phi}_T)\right)\\ &\leq 2\sup_{\phi\in\Phi} |\mathbb{E}_T l_t(\phi) - \mathbb{E}l_t(\phi)| + \mathrm{O_{a.s.}}(T^{-1}),\end{aligned} \tag{A.6}$$

where the last line follows by Lemma A.1.1 as $\mathcal{L}_T(\phi) = \mathcal{L}(\phi) + \mathrm{O_{a.s.}}(T^{-1})$ uniformly over $\phi \in \Phi$ and $\hat{\phi}_T$ is a maximizer of $\phi \mapsto \tilde{\mathcal{L}}_T(\phi)$, so that $\tilde{\mathcal{L}}_T(\hat{\phi}_T) \geq \tilde{\mathcal{L}}_T(\phi_0)$. Under Assumptions A.1.1 and A.1.2, Proposition A.2.1 shows that $(f_t, x_t, u_t)_{t\in\mathbf{Z}_+}$ is stationary and ergodic. Then $(l_t(\phi))_{t\in\mathbf{Z}_+}$ is stationary and ergodic as a measurable transformation of a stationary and ergodic process $(f_t, x_t, u_t)_{t\in\mathbf{Z}_+}$.
Next, we note that it follows from the mean-value theorem that

$$|\ln(1 + e^x)| \leq \ln(2) + |x|.$$

Then

$$\mathbb{E}\left[\sup_{\phi\in\Phi}|l_{i,t}(\phi)|\right] \leq \mathbb{E}\left[\sup_{\phi\in\Phi} 2\left|x_{i,t}^\top\beta + c_i^\top f_t(\phi)\right| + \ln(2)\right],$$

which is finite by Proposition A.2.1 and restrictions on the parameter space in Assumption A.1.3 (i). This observation together with stationarity and ergodicity of $(l_t(\phi_0))_{t\in\mathbf{Z}_+}$ are sufficient for the uniform law of large numbers

$$\sup_{\phi\in\Phi}|\mathbb{E}_T l_t(\phi) - \mathbb{E}l_t(\phi)| \xrightarrow{\text{a.s.}} 0.$$

Combining this with (A.6), we obtain

$$\mathbb{E}l_t(\hat{\phi}_T) \xrightarrow{\text{a.s.}} \mathbb{E}l_t(\phi_0).$$

Under Assumption A.1.3 (iii), this implies that

$$\hat{\phi}_T \xrightarrow{\text{a.s.}} \phi_0.$$

□

To obtain the asymptotic distribution we need the following two auxiliary results.

Lemma A.1.3. *We have*

$$\left\|\frac{\partial^2}{\partial\phi_r\partial\phi_s}s_t(f_t(\phi))\right\| \leq \sum_{j,k}\left|\frac{1}{N_j^k}\sum_{i=1}^n \mathbb{1}\{c_i^k = \iota_j^k\}\pi''_{i,t}(\phi)\xi_{i,t,r}(\phi)\xi_{i,t,s}(\phi) + \pi'_{i,t}(\phi)\frac{\partial}{\partial\phi_r}\xi_{i,t,s}(\phi)\right|,$$

where $\xi_{i,t,r}(\phi) = \frac{\partial}{\partial\phi_r}\left(x_{i,t}^\top\beta + c_i^\top f_t(\phi)\right)$.

Proof. The inequality follows by elementary computations

$$\begin{aligned}
\left\|\frac{\partial^2}{\partial\phi_r\partial\phi_s}s_t(f_t(\phi))\right\|^2 &= \sum_{j,k}\left|\frac{\partial^2}{\partial\phi_r\partial\phi_s}\frac{1}{N_j^k}\sum_{i=1}^n \pi_{i,t}(\phi)\mathbb{1}\{c_i^k = \iota_j^k\}\right|^2 \\
&= \sum_{j,k}\left|\frac{\partial}{\partial\phi_r}\frac{1}{N_j^k}\sum_{i=1}^n \pi'_{i,t}(\phi)\xi_{i,t,s}(\phi)\mathbb{1}\{c_i^k = \iota_j^k\}\right|^2 \\
&= \sum_{j,k}\left|\frac{1}{N_j^k}\sum_{i=1}^n \mathbb{1}\{c_i^k = \iota_j^k\}\left(\pi''_{i,t}(\phi)\xi_{i,t,r}(\phi)\xi_{i,t,s}(\phi) + \pi'_{i,t}(\phi)\frac{\partial}{\partial\phi_r}\xi_{i,t,s}(\phi)\right)\right|^2
\end{aligned}$$

and the inequality $\|x\|_2 \leq \|x\|_1, \forall x \in \mathbf{R}^d$. □

Lemma A.1.4. *Suppose that Assumptions A.1.1, A.1.2, and A.1.3 are satisfied. Then*

$$\mathbb{E}\left[\sup_{\phi\in\Phi}\left\|\frac{\partial^2}{\partial\phi\partial\phi^\top}c_i^\top f_t(\phi)\right\|\right] < \infty.$$

Proof. It is sufficient to show that the expected value of the supremum of the absolute values of all entries of the Hessian matrix is a finite constant. To that end, we need to consider all possible combinations of partial derivatives with respect to $(\alpha, \beta, \delta, \theta)$. There are 10 such derivatives and for the brevity of presentation, we illustrate how to bound two such derivatives. The proof for other partial derivatives is similar.

First, note that by the Cauchy-Schwartz inequality

$$\left|\frac{\partial^2}{\partial\phi_r\partial\phi_s}c_i^\top f_t(\phi)\right| \leq d^{1/2}\left\|\frac{\partial^2}{\partial\phi_r\partial\phi_s}f_t(\phi)\right\|.$$

We focus on the second derivative with respect to different coordinates of α. Put $\alpha_r = (\mathrm{vec}(\alpha))_r$. Then for some j,k dependent on s, we have

$$\begin{aligned}\left\|\frac{\partial^2}{\partial\alpha_r\partial\alpha_s}f_t(\phi)\right\| &\leq \left\|\theta\frac{\partial^2}{\partial\alpha_r\partial\alpha_s}f_{t-1}(\phi)\right\| + \left\|\alpha\frac{\partial^2}{\partial\alpha_r\partial\alpha_s}s_{t-1}(f_{t-1}(\phi))\right\| + 2\max_{r,j,k}\left|\frac{\partial}{\partial\alpha_r}s_{t-1,j,k}(f_{t-1}(\phi))\right| \\ &\leq \rho(\theta)\left\|\frac{\partial^2}{\partial\alpha_r\partial\alpha_s}f_{t-1}(\phi)\right\| + \rho(\alpha)\left\|\frac{\partial^2}{\partial\alpha_r\partial\alpha_s}s_{t-1}(f_{t-1}(\phi))\right\| + \frac{d^{1/2}}{2}\left\|J_\alpha f_{t-1}(\phi)\right\|,\end{aligned} \tag{A.7}$$

where and we used the following estimate that follows by Hölder's inequality

$$\begin{aligned}\max_{r,j,k}\left|\frac{\partial}{\partial\alpha_r}s_{t-1,j,k}(f_{t-1}(\phi))\right| &= \max_{r,j,k}\left|\frac{1}{N_j^k}\sum_{i=1}^n \pi'_{i,t-1}(\phi)\mathbb{1}\{c_i^k = \iota_j^k\}\frac{\partial c_i^\top f_{t-1}(\phi)}{\partial\alpha_r}\right| \\ &\leq \frac{1}{4}\max_r\left\|\frac{\partial}{\partial\alpha_r}f_{t-1}(\phi)\right\|_1 \\ &= \frac{1}{4}\left\|J_\alpha f_{t-1}(\phi)\right\|_1 \\ &\leq \frac{d^{1/2}}{4}\left\|J_\alpha f_{t-1}(\phi)\right\|,\end{aligned}$$

where $\|.\|_1$ denotes the l^1 norm of a vector in the second line and the induced matrix norm in the third line, i.e. for a matrix A with entries $(a_{i,j})$, we have $\|A\|_1 = \max_j\sum_i|a_{i,j}|$. Next we apply Lemma A.1.3 to bound the second term in the upper bound in (A.7). To that end, we note that by Hölder's inequality

$$|\xi_{i,t,r}(\phi)| = \left|c_i^\top\frac{\partial}{\partial\alpha_r}f_t(\phi)\right| \leq \left\|\frac{\partial}{\partial\alpha_r}f_t(\phi)\right\|_1 \leq d^{1/2}\left\|J_\alpha f_t(\phi)\right\|$$

and

$$\left|\frac{\partial}{\partial\alpha_r}\xi_{i,t,s}(\phi)\right| = \left|c_i^\top\frac{\partial^2}{\partial\alpha_r\partial\alpha_s}f_t(\phi)\right| \leq d^{1/2}\left\|\frac{\partial^2}{\partial\alpha_r\partial\alpha_s}f_t(\phi)\right\|.$$

Then by Lemma A.1.3

$$\left\|\frac{\partial^2}{\partial\alpha_r\partial\alpha_s}s_{t-1}(f_{t-1}(\phi))\right\| \leq 0.01d^{3/2}\|J_\alpha f_{t-1}(\phi)\|^2 + \frac{d}{4}\left\|\frac{\partial^2}{\partial\alpha_r\partial\alpha_s}f_{t-1}(\phi)\right\|.$$

Combining all estimates together with (A.7) under Assumption A.1.3 (i), we obtain

$$\left\|\frac{\partial^2}{\partial\alpha_r\partial\alpha_s}f_t(\phi)\right\| \leq c\left\|\frac{\partial^2}{\partial\alpha_r\partial\alpha_s}f_{t-1}(\phi)\right\| + \frac{d^{1/2}}{2}\|J_\alpha f_{t-1}(\phi)\| + 0.01d^{3/2}\rho(\alpha)\|J_\alpha f_{t-1}(\phi)\|^2.$$

In the proof of Lemma A.1.2 we show that $\|J_\alpha f_t(\phi)\| \leq \frac{d}{1-c}$. Therefore,

$$\left\|\frac{\partial^2}{\partial\alpha_r\partial\alpha_s}f_t(\phi)\right\| \leq c\left\|\frac{\partial^2}{\partial\alpha_r\partial\alpha_s}f_{t-1}(\phi)\right\| + O(1)$$

and

$$\sup_{\phi\in\Phi}\left\|\frac{\partial^2}{\partial\alpha_r\partial\alpha_s}f_t(\phi)\right\| = O_{\text{a.s.}}(1).$$

Next we consider the mixed derivatives with respect to α_s and β_r. For some (j,k) depending on s, we have

$$\begin{aligned}\left\|\frac{\partial^2}{\partial\beta_r\partial\alpha_s}f_t(\phi)\right\| &\le \left\|\theta\frac{\partial^2}{\partial\beta_r\partial\alpha_s}f_{t-1}(\phi)\right\| + \left\|\alpha\frac{\partial^2}{\partial\beta_r\partial\alpha_s}s_{t-1}(f_{t-1}(\phi))\right\| + \left|\frac{\partial}{\partial\beta_r}s_{t-1,j,k}(f_{t-1}(\phi))\right| \\ &\le \rho(\theta)\left\|\frac{\partial^2}{\partial\beta_r\partial\alpha_s}f_{t-1}(\phi)\right\| + \rho(\alpha)\left\|\frac{\partial^2}{\partial\beta_r\partial\alpha_s}s_{t-1}(f_{t-1}(\phi))\right\| + \left|\frac{\partial}{\partial\beta_r}s_{t-1,j,k}(f_{t-1}(\phi))\right|.\end{aligned} \tag{A.8}$$

We bound the third term applying the Hölder's inequality

$$\begin{aligned}\left|\frac{\partial}{\partial\beta_r}s_{t-1,j,k}(f_{t-1}(\phi))\right| &= \left|\frac{1}{N_j^k}\sum_{i=1}^n \mathbb{1}\{c_i^k=\iota_j^k\}\pi'_{i,t-1}(\phi)\left\{x_{i,t-1,r}+\frac{\partial}{\partial\beta_r}c_i^\top f_{t-1}\right\}\right| \\ &\le \frac{1}{4}\frac{1}{N_j^k}\sum_{i=1}^n \mathbb{1}\{c_i^k=\iota_j^k\}\|x_{i,t-1}\| + \frac{1}{4}\left\|\frac{\partial}{\partial\beta_r}f_{t-1}(\phi)\right\|_1 \\ &\le \frac{1}{4}\frac{1}{N_j^k}\sum_{i=1}^n \mathbb{1}\{c_i^k=\iota_j^k\}\|x_{i,t-1}\| + \frac{1}{4}\left\|J_\beta f_{t-1}(\phi)\right\|_1,\end{aligned}$$

By (A.5) in the proof of Lemma A.1.2

$$\mathbb{E}\left[\sup_{\phi\in\Phi}\|J_\beta f_{t-1}(\phi)\|\right] = O(1).$$

This shows that

$$\mathbb{E}\left[\sup_{\phi\in\Phi}\left|\frac{\partial}{\partial\beta_r}s_{t-1,j,k}(f_{t-1}(\phi))\right|\right] = O(1).$$

Now we would like to bound the second term in the (A.8) applying Lemma A.1.3. To that end, note that now

$$\begin{aligned}|\xi_{i,t,r}(\phi)| &= \left|x_{i,t,r}+c_i^\top\frac{\partial}{\partial\beta_r}f_t(\phi)\right| \le \|x_{i,t}\| + \left\|\frac{\partial}{\partial\beta_r}f_t(\phi)\right\|_1 \le \|x_{i,t}\| + d^{1/2}\|J_\beta f_t(\phi)\| \\ |\xi_{i,t,s}(\phi)| &= \left|c_i^\top\frac{\partial}{\partial\alpha_s}f_t(\phi)\right| \le \left\|\frac{\partial}{\partial\alpha_s}f_t(\phi)\right\|_1 \le d^{1/2}\|J_\alpha f_t(\phi)\|\end{aligned}$$

and

$$\left|\frac{\partial}{\partial\beta_r}\xi_{i,t,s}(\phi)\right| \le \left|c_i^\top\frac{\partial^2}{\partial\beta_r\partial\alpha_s}f_t(\phi)\right| \le d^{1/2}\left\|\frac{\partial^2}{\partial\beta_r\partial\alpha_s}f_t(\phi)\right\|.$$

Then by Lemma A.1.3

$$\begin{aligned}\left\|\frac{\partial^2}{\partial\beta_r\partial\alpha_s}s_t(f_t(\phi))\right\| &\le \sum_{j,k}\frac{1}{N_j^k}\sum_{i=1}^n \mathbb{1}\{c_i^k=\iota_j^k\}\Bigg|0.01d^{1/2}\|J_\alpha f_t(\phi)\|\left(\|x_{i,t}\|+d^{1/2}\|J_\beta f_t(\phi)\|\right) \\ &\qquad + \frac{d^{1/2}}{4}\left\|\frac{\partial^2}{\partial\beta_r\partial\alpha_s}f_t(\phi)\right\|\Bigg| \\ &\le 0.01\|J_\alpha f_t(\phi)\|\left(\sum_{j,k}\frac{1}{N_j^k}\sum_{i=1}^n \mathbb{1}\{c_i^k=\iota_j^k\}\|x_{i,t}\| + d\|J_\beta f_t(\phi)\|\right) \\ &\qquad + \frac{d}{4}\left\|\frac{\partial^2}{\partial\beta_r\partial\alpha_s}f_t(\phi)\right\|.\end{aligned}$$

Using uniform bounds on Jacobians obtained in the proof of Lemma A.1.2, we have

$$\left\| \frac{\partial^2}{\partial\beta_r\partial\alpha_s} s_{t-1}(f_{t-1}(\phi)) \right\| \le \frac{d}{4} \left\| \frac{\partial^2}{\partial\beta_r\partial\alpha_s} f_{t-1}(\phi) \right\| + \zeta(\phi),$$

where $\mathbb{E}\sup_{\phi\in\Phi}|\zeta(\phi)| = O(1)$ does not depend on t. Combining all estimates with (A.8), we obtain

$$\left\| \frac{\partial^2}{\partial\beta_r\partial\alpha_s} f_t(\phi) \right\| \le c \left\| \frac{\partial^2}{\partial\beta_r\partial\alpha_s} f_{t-1}(\phi) \right\| + \tilde{\zeta}(\phi),$$

for some ζ such that $\mathbb{E}\left[\sup_{\phi\in\Phi}|\zeta(\phi)|\right] = O(1)$. Assumption A.1.3 (i) allows to conclude that

$$\mathbb{E}\left[\sup_{\phi\in\Phi}\left\| \frac{\partial^2}{\partial\beta_r\partial\alpha_s} f_t(\phi) \right\|\right] = O(1).$$

□

We now prove Theorem 1.

Proof. By Lemma A.1.1 and the mean-value expansion of FOCs to the QMLE optimization problem

$$\begin{aligned}\sqrt{Tn}(\hat{\phi}_T - \phi_0) = &- \left(\frac{1}{T}\sum_{t=1}^T \nabla^2 l_t(\bar{\phi})\right)^{-1} \frac{\sqrt{n}}{\sqrt{T}} \sum_{t=1}^T \nabla l_t(\phi_0) \\ &+ \left(\frac{1}{T}\sum_{t=1}^T \nabla^2 l_t(\bar{\phi})\right)^{-1} \sqrt{Tn}\left(\nabla\mathcal{L}_T(\hat{\phi}_T) - \nabla\tilde{\mathcal{L}}_T(\hat{\phi}_T)\right)\end{aligned}$$

for some $\bar{\phi}$ on the line segment with endpoints $\hat{\phi}_T$ and ϕ_0. Under Assumptions A.1.1 and A.1.2, Proposition A.2.1 shows that $(f_t, x_t, u_t)_{t\in\mathbf{Z}_+}$ is geometrically ergodic and so stationary and ergodic if it is initiated from the stationary distribution. By Davydov (1973) for ergodic Markov chain with stationary measure π, the β-mixing coefficient equals to the expected value of the total variation norm with respect to the stationary distribution π

$$\beta_t = \int \|P^t(z,.) - \pi\|_{TV}\pi(\mathrm{d}z),$$

see Appendix A.2 for more details. Therefore, the geometrically ergodic Markov chain $(f_t, x_t, u_t)_{t\in\mathbf{Z}_+}$ is β-mixing with mixing coefficients $\beta_t = O(c^t)$ for some $c \in (0,1)$. Since α-mixing coefficients are bounded by β-mixing coefficients, this implies that $(\nabla l_t(\phi_0))_{t\in\mathbf{Z}_+}$ is α-mixing with the same rate, as a measurable function of a β-mixing process.
In order to apply the mixing CLT we verify that score has finite moment of order $2+\epsilon$ for some $\epsilon > 0$. Let $\xi_{i,t}$ be as in equation (A.2). It follows from the proof of Lemma A.1.2, that the norm of $\xi_{i,t}$ is bounded by the norm of $x_{i,t}$ up to the constant. So $\mathbb{E}\|\xi_{i,t}\|^{2+\epsilon} < \infty$ under Assumption A.1.2 (iii). For a random vector ξ, let $\|\xi\|_{2+\epsilon} = (\mathbb{E}\|\xi\|^{2+\epsilon})^{1/(2+\epsilon)}$ be its $L_{2+\epsilon}$ norm. Then by Minkowski inequality

$$\begin{aligned}\|\nabla l_t(\phi_0)\|_{2+\epsilon} &\le \frac{1}{n}\sum_{i=1}^n |y_{i,t} - \pi_{i,t}(\phi_0, f_t)| \, \|\xi_{i,t}(\phi_0)\|_{2+\epsilon} \\ &\le \|\xi_{i,t}(\phi_0)\|_{2+\epsilon} \\ &\le \|x_{i,t}\|_{2+\epsilon} + \left\|\nabla c_i^\top f_t(\phi_0)\right\|_{2+\epsilon},\end{aligned}$$

which is finite due to the Proposition A.2.1 and Assumption A.1.3 (iii). By the CLT for α-mixing processes, e.g. see Rio (2017)

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\nabla l_t(\phi_0) \xrightarrow{\text{d}} N(0,V),$$

where $V = \lim_{T\to\infty} \operatorname{Var}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\nabla l_t(\phi_0)\right)$ is the long-run variance.
Now we would like to apply the uniform law of large numbers to the Hessian for which we need the moment condition

$$\mathbb{E}\left[\sup_{\phi\in\Phi}\left\|\nabla^2 l_t(\phi)\right\|\right] < \infty.$$

To that end, note that

$$\left\|\nabla^2 l_t(\phi)\right\| \leq \frac{1}{n}\sum_{i=1}^{n}\|\nabla^2 l_{i,t}(\phi)\|$$

and that

$$\begin{aligned}
\left\|\nabla^2 l_{i,t}(\phi)\right\| &\leq \left\|\frac{\partial \xi_{i,t}(\phi)}{\partial\phi^\top}\right\| + \left\|\xi_{i,t}(\phi)\xi_{i,t}(\phi)^\top\right\| \\
&= \left\|\frac{\partial^2}{\partial\phi\partial\phi^\top}c_i^\top f_t(\phi)\right\| + \left\|\xi_{i,t}(\phi)\right\|_\infty \left\|\xi_{i,t}(\phi)\right\|_1 \\
&= \left\|\frac{\partial^2}{\partial\phi\partial\phi^\top}c_i^\top f_t(\phi)\right\| + d^{1/2}\left\|\xi_{i,t}(\phi)\right\|^2 .
\end{aligned}$$

It follows from the proof of Lemma A.1.2 that

$$\mathbb{E}\left[\sup_{\phi\in\Phi}\|\xi_{i,t}(\phi)\|^2\right] = O(1).$$

Therefore, it is sufficient only to check that

$$\mathbb{E}\left[\sup_{\phi\in\Phi}\left\|\frac{\partial^2}{\partial\phi\partial\phi^\top}c_i^\top f_t(\phi)\right\|\right] < \infty,$$

which follows by Lemma A.1.4. So under stationarity and ergodicity of $(\nabla^2 l_t(\phi))_{t\in\mathbf{Z}_+}$ by the uniform law of large numbers

$$\sup_{\phi\in\Phi}\left\|\frac{1}{T}\sum_{t=1}^{T}\nabla^2 l_t(\phi) - J(\phi)\right\| \xrightarrow{\text{a.s.}} 0,$$

where $J(\phi) = \mathbb{E}\left[\nabla^2 l_t(\phi)\right]$.
The result follows by Theorem A.2, the continuous mapping theorem, and Slutsky's lemma.

□

A.2 Geometric ergodicity

There are many different strategies to verify that the stochastic process of interest is stationary and ergodic. For nonlinear autoregressive processes, a popular approach is to show that the function describing nonlinearities satisfies certain stochastic contracting Lipschitz properties, see Bougerol (1993). However, this does not apply in our case, because this function

for the factor process is discontinuous in factors. Moreover, stationarity and ergodicity alone are not sufficient for the central limit theorem, so we also need mixing[15].
An alternative approach is based on recognizing that the nonlinear autoregressive process is Markov and exploiting results available for Markov chains, Meyn and Tweedie (2012). In this paper, we follow this route. Let $\mathcal{B}(\mathbf{R}^d)$ be a collection of Borel sets. The Markov chain on $(\mathbf{R}^d, \mathcal{B}(\mathbf{R}^d))$ is geometrically ergodic if there exists a constant $c \in (0,1)$ and a measure π on $\mathbf{R}^d$ such that

$$c^{-t}\|P^t(z,.) - \pi\|_{TV} \xrightarrow[t\to\infty]{} 0, \qquad \forall z \in \mathbf{R}^d,$$

where $P^t(z,A) = \Pr(X_n \in A | X_0 = z)$ is a transition kernel and $\|.\|_{TV}$ is a total variation norm. Geometric ergodicity tells us the speed at which the convergence to the stationary distribution takes place for arbitrary initial value $z \in \mathbf{R}^d$. It follows from Davydov (1973) that geometrically ergodic Markov chain is β-mixing, which in turn is sufficient for the CLT under mild moment conditions.
The nonlinear autoregressive equation for factors is

$$f_t = \delta + \theta f_{t-1} + \alpha s_{t-1}(\phi, f_{t-1}).$$

In the following proposition we show that the Markov chain $(f_t)_{t\in\mathbf{Z}_+}$ satisfying this equation is geometrically ergodic and that it is bounded in its stationary regime. As a result f_t will have moments of all orders.

Proposition A.2.1. *Suppose that assumptions A.1.1, A.1.2, and A.1.3 are satisfied. Then* $(f_t, x_t, u_t)_{t\in\mathbf{Z}_+}$ *is geometrically ergodic and* $\|f_t\| \leq \frac{\|\delta\|+d^{1/2}\rho(\alpha)}{1-\rho(\theta)}$ *for any* $\phi \in \Phi$.

Proof. Let $\mathcal{F}_t = \sigma(f_s : \; s \leq t)$ be filtration generated by the past of the factor process. Note that $f_t = \Psi(f_{t-1}, \epsilon_{t-1})$ for some function Ψ and $\epsilon_{t-1} = (x_{t-1}, u_{t-1})$. Under assumption A.1.1

$$f_{t+1}|\mathcal{F}_t =_d \Psi(f_t, \epsilon_t)|\mathcal{F}_t =_d \Psi(f_t, \epsilon_t)|f_t =_d f_{t+1}|f_t.$$

This shows that (f_t) is a Markov chain. By (Feigin and Tweedie, 1985, Theorem 1) the Markov chain is geometrically ergodic if it is Feller, irreducible, and satisfied the geometric drift condition. For any continuous and bounded function $\psi : \mathbf{R}^k \to \mathbf{R}$

$$\mathbb{E}[\psi(f_{t+1})|f_t = z] = \mathbb{E}[\psi(\delta + \theta z + \alpha s_t(\phi, z))]$$

is bounded and continuous function of z. This follows after the change of variables by the Lebesgue's dominated convergence theorem due to the fact that u_t has continuous Lebesgue density. Therefore, the Markov chain is Feller.
Now we show that the process is irreducible, that is for any Borel set $A \subset \mathbf{R}^d$

$$\lambda_R(A) > 0 \implies \forall z \in \mathbf{R}^d \,, \exists t \in \mathbf{Z}_+ \quad \text{s.t.} \quad \Pr(f_t \in A | f_0 = z) > 0,$$

where $R \subset \mathbf{R}^d$ is a compact set of positive Lebesgue measure to be specified below and λ_R is the restriction of the Lebesgue measure to R. Iterating backwards, conditionally on $f_0 = z$ we obtain

$$f_t = (I-\theta)^{-1}(I-\theta^t)\delta + \theta^t z + \sum_{j=1}^{t} \theta^{j-1}\alpha s_{t-j}(\phi, f_{t-j})$$

[15] For correctly specified dynamic models, the score is a martingale difference sequence. Here, we consider the possibility of model misspecification and use QMLE, which is why we need mixing conditions

Let F be the CDF of $u_{i,t}$. Then $U_{i,t} = F(u_{i,t})$ are i.i.d. uniformly distributed on $[0,1]$. Take $a, b \in (0,1), a < b$ and consider the following event

$$B_t = \bigcap_{s=0}^{t} \bigcap_{i=1}^{n} \{a \leq \pi_{i,s} \leq b\} \cap \left\{U_{i,s} > F \circ \pi^{-1}(b)\right\},$$

where we denote $\pi_{i,s} = \pi_{i,s}(\phi)$ and π is logistic CDF. Note that on this event $y_{i,s} = 0$, $\forall i = 1, \dots, n, s = 1, \dots, t$. Since $U_{i,t}$ are i.i.d. $U[0,1]$ and independent from all other variables, $\Pr(B_t) > 0$ if we show that $(\pi_{1,0}, \dots, \pi_{n,t})$ have positive Lebesgue density on $[a,b]^{n(t+1)}$ conditionally on $f_0 = z$. To that end, note that $f_t = \Delta_t(x_0^{t-1}, u_0^{t-1})$ for some function Δ_t, where $x_0^{t-1} = (x_s^\top \beta)_{s=0}^{t-1}$ and $u_0^{t-1} = (u_s)_{s=0}^{t-1}$. Then for $i = 1, \dots, n$ and $s = 1, \dots, t$

$$\pi_{i,s} = \pi\left(x_{i,s}^\top \beta + c_i^\top \Delta\left(x_0^{s-1}, u_0^{s-1}\right)\right)$$

and

$$\pi_{i,0} = \pi\left(x_{i,0}^\top \beta + c_i^\top z\right).$$

Since the process (u_t) is independent from (x_t)

$$\begin{aligned}
&\Pr\left(\pi_{i,s} \leq w_{i,s}, 1 \leq i \leq n, 0 \leq s \leq t\right) \\
&= \Pr\left(x_{i,s}^\top \beta \leq \pi^{-1}(w_{i,s}) - \Delta\left(x_0^{s-1}, u_0^{s-1}\right), 1 \leq i \leq n, 0 \leq s \leq t\right) \\
&= \int \int_{C_0} \cdots \int_{C_t} f_{x_0^t}(x_0, \dots, x_t) \mathrm{d}x_1 \dots \mathrm{d}x_t f_{u_0^{t-1}}(u) \mathrm{d}u,
\end{aligned}$$

where

$$\begin{aligned}
C_s &= \prod_{i=1}^{n} \left(-\infty, \pi^{-1}(w_{i,s}) - \Delta\left(x_0^{s-1}, u_0^{s-1}\right)\right] \\
C_0 &= \prod_{i=1}^{n} \left(-\infty, \pi^{-1}(w_{i,0}) - c_i^\top z\right].
\end{aligned}$$

Applying the differential operator $\frac{\partial^{n(t+1)}}{\partial w_{1,0} \dots \partial w_{n,0} \dots \partial w_{1,t} \dots \partial w_{n,t}}$ to the above expression for some sequence of bounded functions $\delta_{i,s}(u, w)$ we obtain the joint density function

$$f_\pi(w_{1,0}, \dots, w_{n,t}) = C(w) \int f_{x_0^t}\left(\pi^{-1}(w_{1,0}) - c_1^\top z, \dots, \pi^{-1}(w_{n,t}) - \delta_{n,t-1}(u, w)\right) f_{u_0^{t-1}}(u) \mathrm{d}u,$$

where $C(w) = \prod_{i=1}^{n} \prod_{s=0}^{t} \frac{1}{\pi'(\pi^{-1}(w_{i,s}))}$. This density is strictly positive on $[a,b]^{n(t+1)}$, under Assumption A.1.2 (i)-(ii), whence, $\Pr(B_t) > 0$. Now on the event B_t

$$f_t = (I - \theta)^{-1}(I - \theta^t)\delta + \theta^t z - \Pi,$$

where

$$\begin{aligned}
\Pi &= \left(\alpha \bar{\pi}_t + \theta \alpha \bar{\pi}_{t-1} + \dots + \theta^{t-1} \alpha \bar{\pi}_0\right) \\
\bar{\pi}_t &= \left(\frac{1}{N_j^k} \sum_{i=1}^{N_j^k} \pi_{i,t} \mathbb{1}\left\{c_i^k = \iota_j^k\right\}\right)_{\substack{j=1,\dots,d_k \\ k=1,\dots,d}}.
\end{aligned}$$

For a matrix A of dimension $d \times d$, let $M_A = \max_{x \in [a,b]^d} Ax$ and $m_A = \min_{x \in [a,b]^d} Ax$ be vectors of coordinate-wise maximum and minimum. For example, if $A^\top = (a_1^\top, \dots, a_d^\top)^\top$, then $(M_A)_j = \max_{x\in[a,b]^d} a_j^\top x$ and $(m_A)_j = \min_{x\in[a,b]^d} a_j^\top x$. Then since $(\bar\pi_0,\dots,\bar\pi_t)$ has strictly positive density on $[a,b]^{d(t+1)}$, Π has strictly positive density on the rectangle $[\underline{\Pi}_t, \overline{\Pi}_t]$ with

$$\begin{aligned}\underline{\Pi}_t &= m_\alpha + m_{\theta\alpha} + \cdots + m_{\theta^{t-1}\alpha}\\ \overline{\Pi}_t &= M_\alpha + M_{\theta\alpha} + \cdots + M_{\theta^{t-1}\alpha}.\end{aligned}$$

Therefore, on the event B_t, f_t has strictly positive density on the rectangle

$$R_t = \left\{x\in\mathbf{R}^d:\ (I-\theta)^{-1}(I-\theta^t)\delta + \theta^t z + \underline{\Pi}_t \le x \le (I-\theta)^{-1}(I-\theta^t)\delta + \theta^t z + \overline{\Pi}_t\right\}.$$

Put

$$R = \left\{x\in\mathbf{R}^d:\ (I-\theta)^{-1}\delta + \underline{\Pi} \le x \le (I-\theta)^{-1}\delta + \overline{\Pi}\right\},$$

where $\underline{\Pi} = \sum_{t=0}^\infty m_{\theta^t\alpha}$ and $\overline{\Pi} = \sum_{t=0}^\infty M_{\theta^t\alpha}$ converge under Assumption A.1.3. Let λ_R be restriction of Lebesgue measure to R. Then for any measurable $A\subset\mathbf{R}^d$ with $\lambda_R(A)>0$, there exists $t\in\mathbf{Z}_+$ such that $\lambda(A\cap R_t)>0$, whence

$$\Pr(f_t\in A|f_0=z) \ge \Pr(f_t\in A|B_t\cap\{f_0=z\})\Pr(B_t) = \int_A f_{f_t}(x)\mathrm{d}\lambda(x)\Pr(B_t) > 0.$$

Now we verify the geometric drift condition. Let A be a compact set such that $\lambda_R(A)>0$. We need to show that there exists a function $V:\mathbf{R}^d\to\mathbf{R}$ and a constant $L\in(0,1)$ such that

$$\begin{aligned} V(z) &\ge 1, \qquad \forall z\in A\\ \mathbb{E}[V(f_{t+1})|f_t=z] &\le LV(z), \qquad \forall z\in A^c.\end{aligned}$$

Take $V(z) = 1+\|z\|$. Then $V(z)\ge 1, \forall z\in A$, where A is arbitrary subset of $\mathbf{R}^d$ and

$$\begin{aligned}\mathbb{E}[V(f_{t+1})|f_t=z] &\le 1+\rho(\theta)\|z\| + \mathbb{E}\|\delta+\alpha s_t(\phi,z)\|\\ &\le LV(z),\end{aligned}$$

for any $L\in(\rho(\theta),1)$ and

$$z\in A^c = \left\{z\in\mathbf{R}^d:\ \|z\| > \frac{1-L+\mathbb{E}\|\delta+\alpha s_0(\phi,z)\|}{L-\rho(\theta)}\right\}.$$

Clearly, for this choice of A^c, the set A is compact and $\lambda_R(A)>0$ if we choose L sufficiently close to $\rho(\theta)$. This verifies the geometric drift condition and shows that $(f_t)_{t\in\mathbf{Z}_+}$ is geometrically ergodic. The geometric ergodicity of (f_t,x_t,u_t) follows from the geometric ergodicity of f_t and stationarity of (x_t,u_t).
For the last statement, we invert the linear part of the process

$$f_t = (I-\theta)^{-1}\delta + \sum_{j=0}^\infty \theta^j\alpha s_{t-j-1}.$$

and bound it as follows

$$\begin{aligned}\|f_t\| &\le (1-\rho(\theta))^{-1}\|\delta\| + \rho(\alpha)\sum_{j=0}^\infty \rho(\theta)^j\|s_{t-j-1}\|\\ &\le \frac{\|\delta\| + d^{1/2}\rho(\alpha)}{1-\rho(\theta)}.\end{aligned}$$

□