

# TEMPORAL CORRELATION OF DEFAULTS IN SUBPRIME SECURITIZATION

ERIC HILLEBRAND, AMBAR N. SENGUPTA, AND JUNYUE XU

**ABSTRACT.** The securitization of subprime mortgages in instruments like mortgage-backed securities and collateralized debt obligations is one of the key ingredients to the current financial crisis. During 2007 and 2008, subprime defaults increased sharply, displaying high serial correlation in their arrival. Subprime default events depend on house price changes. We establish a link between the dynamics of house price changes and the dynamics of default rates in the Gaussian copula framework by specifying a time series model for a common risk factor. We show analytically and in simulations that serial correlation propagates from the common risk factor to default rates. We simulate prices of mortgage-backed securities, which are securitized from pools of mortgages using a waterfall structure. We find that subsequent vintages of these securities inherit temporal correlation from the common risk factor. The findings in this paper formalize one important dynamic of the subprime crisis: transmission of the decline in housing prices after 2006 into financial derivatives based on subprime mortgages.

**KEYWORDS:** subprime mortgage, housing prices, mortgage-backed securities, collateralized debt obligations, financial crisis, vintage correlation, serial correlation, time series model, Gaussian copula

**JEL Codes:** C02, E44, G01, G13, G21

## 1. INTRODUCTION

From its beginning in the summer of 2007, the subprime crisis has plunged the world into one of the worst recessions in history. At the center of the crisis is the subprime mortgage market, where lenders provide mortgages to borrowers with poor credit standing. During the crisis, subprime mortgages created at different times have defaulted one after another. The default arrivals of these mortgages were serially correlated. Figure 1, lower panel, shows the time series of serious delinquency rates of subprime mortgages from 2002 to 2009.<sup>1</sup> This series obviously displays very high serial correlation. Defaults of subprime mortgages are closely connected to house price fluctuations, as suggested, among others, by Gorton (2008).<sup>2</sup> Most subprime mortgages are Adjustable-Rate Mortgages (ARM). This means that the interest rate on a subprime mortgage is fixed at a relatively low level for a “teaser” period, usually two to three years, after which it increases substantially. Gorton (2008) points out that the interest rate usually resets to such a high level that it “essentially forces” a mortgage borrower to refinance or default after the teaser period. Therefore, whether the mortgage defaults or not is largely determined by the borrower’s access to refinancing. At the end of the teaser period, if the value of the house is much greater than the outstanding principal of the loan, the borrower is likely to be approved

---

*Date:* March 15, 2010.

<sup>1</sup>By definition of the Mortgage Banker Association, seriously delinquent mortgages refer to mortgages that have either been delinquent for more than 90 days or are in the process of foreclosure.

<sup>2</sup>For example, see also Bajari, Chu, and Park (2008), Daglish (2009), Hayre, Saraf, Young, and Chen (2008).

for a new loan since the house serves as collateral. On the other hand, if the value of the house is less than the outstanding principal of the loan, the borrower is unlikely to be able to refinance and has to default.

Gorton's view is supported by data. Figure 1 displays two-year changes in the Case-Shiller index (upper panel) and subprime ARM serious delinquency rates (lower panel) using quarterly data from 2002 to 2009. It is clear that from 2002 to 2006, subprime delinquency rates declined as home prices climbed steadily. The delinquency series reached its trough around the same time the home price peaked. When the house price index started to drop in 2006, delinquency rates began to increase significantly, which triggered the subprime crisis.

Therefore, the hypothesis that this paper examines is that the dynamics of defaults are inherited from the dynamics of house prices. The aim of this paper is to formalize this relationship using the industry-standard framework of a Gaussian copula, which was routinely used to price derivatives constructed from subprime mortgages. To this end, we introduce the notion of *vintage correlation*, which captures the correlation of default rates in mortgage pools issued at different times. Under certain assumptions, vintage correlation is the same as serial correlation. After showing that changes in a housing index can be regarded as a common risk factor of individual subprime mortgages, we specify a time series model for the common risk factor in the Gaussian copula framework. We show analytically and in simulations that the serial correlation of the common risk factor introduces vintage correlation into default rates of pools of subprime mortgages of subsequent vintages. In this sense, serial correlation propagates from the common risk factor to default rates. In simulations of the price behavior of Mortgage-Backed Securities (MBS) over different cohorts, we find that the price of MBS also exhibits vintage correlation, which is inherited from the common risk factor of individual mortgages.

The main point of this paper is to provide a formal examination of one of the important causes of the current crisis.<sup>3</sup> Vintage correlation in default rates and MBS prices also has implications for asset pricing. To price some derivatives, for example forward starting Collateralized Debt Obligations (CDO), it is necessary to predict default rates of credit assets created at some future time. Knowing the serial correlation of default probabilities can improve the quality of prediction. For risk management in general, some credit asset portfolios may consist of credit derivatives of different cohorts. Vintage correlation of credit asset performance affects these portfolios' risks. For instance, suppose there is a portfolio consisting of two subsequent vintages of the same MBS. If the vintage correlation of the MBS price is close to one, for example, the payoff of the portfolio has a variance almost twice as big as if there were no vintage correlation.

The outline of the paper is as follows. In Section 2, we introduce the concept of vintage correlation and give some examples to provide intuition. We then briefly describe the Gaussian copula model. We show that changes in a house price index can be seen as a common risk factor in the copula framework. Section 3 contains the main analytical results. It shows the link between the serial correlation of the

---

<sup>3</sup>For different perspectives on the causes and effects of the subprime crisis, see also Caballero and Krishnamurthy (2009), Crouhy, Jarrow, and Turnbull (2008), Figlewski (2009), Gorton (2009), Murphy (2008), Reinhart and Rogoff (2008), and Reinhart and Rogoff (2009).

common risk factor and vintage correlation in default rates. Section 4 explores this link in two sets of simulations: First, a series of mortgage pools is simulated to confirm our analytical results. Second, a waterfall structure is simulated to study the propagation of serial correlation in MBS. In Section 5 we summarize the main conclusions.

## 2. MODELING TEMPORAL CORRELATION IN SUBPRIME SECURITIZATION

In this section, we introduce the concept of vintage correlation and give some examples to provide intuition. We outline the Gaussian copula model. We show that changes in a house price index can be seen as a common risk factor in the copula framework.

**DEFINITION 1 (Vintage Correlation).** *Suppose we have a pool of mortgages created at each time  $v = 1, 2, \dots, V$ . Denote the default rates of each vintage observed at a fixed time  $T > V$  as  $p_1, p_2, \dots, p_V$ , respectively. We define vintage correlation  $\phi_j := \text{Corr}(p_1, p_j)$  for  $j = 2, 3, \dots, V$  as the default correlation between the  $j$ -th vintage and the first vintage.*

As an example of vintage correlation, consider wines of different vintages. Suppose there are several wine producers that have produced wines of ten vintages from 2011 to 2020. The wines are packaged according to vintages and producers, that is, one box contains one vintage by one producer. In the year 2022, all boxes are opened and the percentage of wines that have gone bad is obtained for each box. Consider the correlation of fractions of bad wines between the first vintage and subsequent vintages. This correlation is what we call vintage correlation.

The definition of vintage correlation can be extended easily to the case where the base vintage is not the first vintage but any one of the other vintages. Obviously, vintage correlation is very similar to serial correlation. There are two main differences. First, the consideration is at a specific time in the future. Second, in calculating the correlation between any two vintages, the expected values are averages over the cross-section. That is, in the wine example, expected values are averages over producers. In mortgage pools, they are averages over different mortgage pools. Only if we assume the same stochastic structure for the cross-section and for the time series of default rates, vintage correlation and serial correlation are equivalent. We do not have to make this assumption to obtain our main results. Making this assumption, however, does not invalidate any of the results either. Therefore, we use the terms “vintage correlation” and “serial correlation” interchangeably in our paper.

To model vintage correlation in subprime securitization, we resort to the copula approach. The Gaussian copula approach is widely used in industry to model default correlation across names. A copula is a function that takes the marginal distribution functions of a set of variables as arguments and returns the joint distribution of the variables. Thus, the copula approach provides a general way to link univariate marginal distribution functions to their multivariate distribution function. This feature makes it very useful for modeling multivariate correlations. Frees and Valdez (1998) explain in detail how to specify a copula, and how to simulate a multivariate distribution once its copula form is known.

The credit industry standard copula model was introduced by Li (2000) and is called default-time (or survival-time) Gaussian copula. This model is applicable to all types of CDO, MBS, and almost

all other credit derivatives that are derived from multiple assets with credit risk. The idea behind Li's model is that each credit asset has a default time (or survival time), after which the mortgage defaults. Instead of modeling the correlation between default events of mortgages, Li proposes a copula approach to capture the joint distribution of default times. A copula in this case takes the marginal distribution of default times and returns their joint distribution.

The literature on credit risk pricing with copulas and other models has grown substantially in recent years and an exhaustive review is beyond the scope of his paper. Bluhm, Overbeck, and Wagner (2002), Schönbucher (2003), Duffie and Singleton (2003), and Lando (2004) are standard monographs. Some modifications of the standard Gaussian copula model are discussed. For example, Servigny and Renault (2002) and Das, Freed, Geng, and Kapadia (2006) provide empirical evidence that asset correlation may be stochastic. Andersen and Sidenius (2005), Hull, Predescu, and White (2009), and Berd, Engle, and Voronov (2007) allow default correlation to vary over time. Copula models using distribution functions other than Gaussians have also been suggested. For example, Andersen, Sidenius, and Basu (2003) and Frey and McNeil (2003) consider the student  $t$ -copula. Schönbucher and Schubert (2001) and Laurent and Gregory (2005) discuss the Clayton copula. The Marshall-Olkin copula has been considered by Lindskog and McNeil (2003) and Giesecke (2003). A recent study by Beare (2010) explicitly addresses the temporal correlation problem in copulas.

There are approaches to model default correlation other than default-time copulas. One method relies on the so-called structural model, which goes back to Merton's (1974) work on pricing corporate debt. An essential point of the structural model is that it links the default event to some observable economic variables. Hull and White (2001) extend the model to a multi-issuer scenario, which can be applied to price corporate debt CDO. It is assumed that a firm defaults if its credit index hits a certain barrier. Therefore, correlation between credit indices determines the correlation of default events. The advantage of a structural model is that it gives economic meaning to underlying variables. Other approaches to CDO pricing are found, for example, in Graziano and Rogers (2009) and in Sidenius, Piterbarg, and Andersen (2008). Burtschell, Gregory, and Laurent (2009) provide a comparison of common CDO pricing models.

In this paper, we adopt Li's (2000) default time copula approach and extend it by adding a time series model for a common risk factor. Each mortgage  $i$  of vintage  $v$  has a default time  $\tau_{v,i}$ , which is a random variable representing the time at which the mortgage defaults. If the mortgage never defaults, this value is infinity. If we assume that the distribution of  $\tau_{v,i}$  is the same across all mortgages of vintage  $v$ , we have

$$F_v(s) = \mathbb{P}[\tau_{v,i} < s], \forall i = 1, 2, \dots, N, \quad (1)$$

where the index  $i$  denotes individual mortgages and the index  $v$  denotes vintages. We assume that  $F_v$  is continuous and strictly increasing. Given this information, for each vintage  $v$  the Gaussian copula approach provides a way to obtain the joint distribution of the  $\tau_{v,i}$  across  $i$ . Generally, a copula is a joint distribution function

$$C(u_1, u_2, \dots, u_N) = \mathbb{P}(U_1 \leq u_1, U_2 \leq u_2, \dots, U_N \leq u_N),$$

where  $u_1, u_2, \dots, u_N$  are  $N$  uniformly distributed random variables that may be correlated. It can be easily verified that the function

$$C [F_1(x_1), F_2(x_2), \dots, F_N(x_N)] = G(x_1, x_2, \dots, x_N) \quad (2)$$

is a multivariate distribution function with marginal distribution functions  $F_1(x_1), F_2(x_2), \dots, F_N(x_N)$ . Sklar (1959) proved the converse. He showed that for an arbitrary multivariate distribution function  $G(x_1, x_2, \dots, x_N)$  with continuous marginal distributions functions  $F_1(x_1), F_2(x_2), \dots, F_N(x_N)$ , there exists a unique  $C$  such that equation (2) holds. Therefore, in the case of default times, there is a  $C_v$  for each vintage  $v$  such that

$$C_v [F_v(\tau_{v,1}), F_v(\tau_{v,2}), \dots, F_v(\tau_{v,N})] = G_v(\tau_{v,1}, \tau_{v,2}, \dots, \tau_{v,N}). \quad (3)$$

The joint distribution function  $G_v$  on the right-hand side of equation (3) is the object we want to obtain. Since we assume  $F_v$  to be continuous and strictly increasing, we can find a standard Gaussian random variable  $X_{v,i}$  such that

$$\Phi(X_{v,i}) = F_v(\tau_{v,i}) \quad \forall v = 1, 2, \dots, V; i = 1, 2, \dots, N, \quad (4)$$

or equivalently,

$$\tau_{v,i} = F_v^{-1}(\Phi(X_{v,i})) \quad \forall v = 1, 2, \dots, V; i = 1, 2, \dots, N, \quad (5)$$

where  $\Phi$  is the standard normal distribution function. To see that this is correct, observe that

$$\begin{aligned} \mathbb{P}[\tau_{v,i} \leq s] &= \mathbb{P}[\Phi(X_{v,i}) \leq F_v(s)] \\ &= \mathbb{P}[X_{v,i} \leq \Phi^{-1}(F_v(s))] \\ &= \Phi[\Phi^{-1}(F_v(s))] \\ &= F_v(s). \end{aligned}$$

Substituting equation (4) into the left-hand side of equation (3), we have

$$C_v [\Phi(X_{v,1}), \Phi(X_{v,2}), \dots, \Phi(X_{v,N})] = G_v(\tau_{v,1}, \tau_{v,2}, \dots, \tau_{v,N}). \quad (6)$$

Since  $\Phi(\cdot)$  is the marginal distribution function for all  $X_{v,i}$ , the left-hand side of equation (6) is equal to the joint distribution of  $X_{v,i}$ . The Gaussian copula approach assumes that this joint distribution has a multivariate normal distribution function  $\Phi_N$ ,

$$G_v(\tau_{v,1}, \tau_{v,2}, \dots, \tau_{v,N}) = \Phi_N(X_{v,1}, X_{v,2}, \dots, X_{v,N}). \quad (7)$$

Thus the joint distribution function of default times  $\tau_{v,i}$  is obtained once the correlation matrix of the  $X_{v,i}$  is known. A standard simplification in practice is to assume that the pairwise correlations between different  $X_{v,i}$  are the same across  $i$ . Suppose that the value of this correlation is  $\rho_v$  for each vintage  $v$ . Consider the following definition

$$X_{v,i} := \sqrt{\rho_v} Z_v + \sqrt{1 - \rho_v} \varepsilon_i \quad \forall i = 1, 2, \dots, N; v = 1, 2, \dots, V, \quad (8)$$

where  $\varepsilon_{v,i}$  are i.i.d. standard Gaussian random variables and  $Z_v$  is a Gaussian random variable independent of the  $\varepsilon_{v,i}$ . It can be shown easily that in each vintage  $v$ , the variables  $X_{v,i}$  defined in this way have the exact joint distribution function  $\Phi_N$ .

Using the information above, for each vintage  $v$ , the Gaussian copula approach obtains the joint distribution function  $G_v$  for default times as follows. First,  $N$  Gaussian random variables  $X_{v,i}$  are generated according to equation (8). Second, from equation (5) a set of  $N$  default times  $\tau_{v,i}$  is obtained, which has the desired joint distribution function  $G_v$ . In equation (8), the common factor  $Z_v$  can be viewed as a latent variable that captures the default risk in the economy, and  $\varepsilon_i$  is the idiosyncratic risk for each mortgage. The variable  $X_{v,i}$  can be viewed as a state variable for each mortgage. The parameter  $\rho_v$  is the correlation between any two individual state variables. It is obvious that the higher the value of  $\rho_v$ , the greater the correlation between the default times of different mortgages.

In pricing derivatives created on subprime mortgages, Monte Carlo simulations are employed to study the default behavior of mortgages by the method described above. In each simulation, the default times for all mortgages are generated with the joint distribution  $G_v$ . Mortgage  $i$  is said to default before time  $T$ , if its simulated default time  $\tau_{v,i}$  is less than  $T$ . The value of  $\rho_v$  can then be calibrated to market data. The market-implied  $\rho_v$  may vary over time. Indeed, this is supported by empirical evidence provided by Servigny and Renault (2002) and Das, Freed, Geng, and Kapadia (2006). This time dependence may capture dynamic correlation between default events not explicitly captured in the default time copula approach. One way to explicitly model the dynamics of defaults is to specify a stochastic process for default correlation. This approach is called stochastic correlation (see for example Andersen and Sidenius (2005) and Hull, Predescu, and White (2009)). Das, Freed, Geng, and Kapadia (2006) propose a model where the default intensity is determined by the state of the economy, which follows a Markov process.

In this paper, we propose a time series model for the common risk factor  $Z_v$  in the copula framework and show that its serial correlation propagates to the default rates. To illustrate the intuition behind our approach, we first give a structural interpretation for the common risk factor  $Z_v$  of subprime ARM.

Assume that we have a pool of  $N$  mortgages  $i = 1, \dots, N$  for each vintage  $v = 1, \dots, V$ . Each individual mortgage within a pool has the same initiation date  $v$  and interest adjustment date  $v' > v$ . Let  $Y_{v,i}$  be the change in the logarithm of the price  $P_{v,i}$  of borrower  $i$ 's (of vintage  $v$ ) house during the teaser period  $[v, v']$ . Consider

$$Y_{v,i} := \log P_{v',i} - \log P_{v,i} = H_v + e_{v,i}, \quad (9)$$

where  $H_v := \log I_{v'} - \log I_v$  is the change in the logarithm of a housing market index  $I_v$ , and  $e_{v,i}$  are i.i.d. normal random variables for all  $i = 1, 2, \dots, N$ , and  $v = 1, 2, \dots, V$ . As outlined in the introduction, default rates of subprime ARM depend on house price changes during the teaser period. If the house price fails to increase substantially or even declines, the mortgage borrower cannot refinance, absent other substantial improvements in income or asset position. They have to default shortly after the interest rate is reset to a high level. We assume that the default, if it happens, occurs at time  $v'$ . Therefore, we assume that a mortgage defaults if and only if  $Y_{v,i} < Y^*$ , where  $Y^*$  is a predetermined

threshold. For example, if we set  $Y^* = 0$ , we are implicitly assuming that if the house price increases, the mortgage borrower is able to refinance. Otherwise, they cannot be approved for a new loan and have to default. Suppose we have a portfolio of  $N$  mortgages, which satisfy the assumptions above. Now, if we assume a form of stationary stochastic process for  $H_v$ , say an ARMA(p,q) process, we can simulate the default rates over time in the portfolio by Monte Carlo simulation.

We can now give a structural interpretation of the common risk factor  $Z_v$  in the Gaussian copula framework. Define

$$Z'_v := \frac{H_v}{\sigma_H}, \quad (10)$$

where  $\sigma_H$  is the unconditional standard deviation of  $H_v$ . Then we have

$$Y_{v,i} = Z'_v \sigma_H + e_{v,i}.$$

Further standardizing  $Y_{v,i}$ , we have

$$\begin{aligned} X'_{v,i} &:= \frac{Y_{v,i}}{\sigma_Y} \\ &= \frac{Z'_v \sigma_H + e_{v,i}}{\sigma_Y} \\ &= \frac{\sigma_H}{\sqrt{\sigma_H^2 + \sigma_e^2}} Z'_v + \frac{\sigma_e}{\sqrt{\sigma_H^2 + \sigma_e^2}} \varepsilon'_{v,i} \end{aligned}$$

where  $\sigma_e$  is the standard deviation of  $e_{v,i}$ , and  $\varepsilon'_{v,i} := e_{v,i}/\sigma_e$ . The third equality follows from the fact that

$$\sigma_Y = \sqrt{\sigma_H^2 + \sigma_e^2}.$$

Define

$$\rho' := \frac{\sigma_H^2}{\sigma_H^2 + \sigma_e^2}.$$

Then

$$X'_{v,i} = \sqrt{\rho'} Z'_v + \sqrt{1 - \rho'} \varepsilon'_{v,i} \quad \forall i = 1, 2, \dots, N; \quad t = 1, 2, \dots, T \quad (11)$$

Note that equation (11) has exactly the same form as equation (8). The default event is defined as  $X'_{v,i} < X^{*'} where$

$$X^{*'} := \frac{Y^*}{\sqrt{\sigma_H^2 + \sigma_e^2}}.$$

Let

$$\tau'_{v,i} := F_v^{-1}(\Phi(X'_{v,i})),$$

and

$$\tau_v^{*'} := F_v^{-1}(\Phi(X^{*'})),$$

then the default event can be defined equivalently as  $\tau'_{v,i} \leq \tau_v^{*'}$ . The comparison between equation (11) and (8) shows that the common risk factor  $Z_v$  in the Gaussian copula model for subprime mortgages can be interpreted as a standardized change in a house price index.

In light of this structural interpretation, the common risk factor  $Z_v$  is very likely to be serially correlated across subsequent vintages. More specifically, we find that  $Z'_v$  is proportional to a moving average of monthly log changes in a housing price index. To see this, let  $v$  be the time of origination and  $v'$  be the end of the teaser period. Then,

$$H_v = \int_v^{v'} d \log I_\tau,$$

where  $I$  is the house price index. For example, if we measure house price index changes quarterly, as in the case of the Case-Shiller housing index, we have

$$H_v = \sum_{\tau \in [v, v']} (\log I_\tau - \log I_{\tau-1}), \quad (12)$$

where the unit of  $\tau$  is a quarter. If we model this index by some random shock arriving each quarter, equation (12) is a moving average process. Therefore, from equation (10) we know that  $Z'_v$  has positive serial correlation.

### 3. MAIN THEOREMS - VINTAGE CORRELATION IN DEFAULT RATES

Since the common risk factor is likely to be serially correlated, we examine the implications for the stochastic properties of mortgage default rates. Most subprime ARM have a teaser period of two years, therefore equation (12) suggests that a two-year house index change can be used as a common risk factor for these mortgages. Figure 1 compares two-year changes in the Case-Shiller index with subprime ARM serious delinquency rates. The two variables are highly and negatively correlated with each other. To study this observation analytically, we specify a time series model for the common risk factor in the Gaussian copula approach. We then determine the relationship between the serial correlation of the default rates and that of the common risk factor.

PROPOSITION 1 (Default Probabilities and Numbers of Defaults). *Let  $k = 1, 2, \dots, N$ ,*

$$X_k = \sqrt{\rho}Z + \sqrt{1 - \rho}\varepsilon_k, \quad \text{and} \quad X'_k = \sqrt{\rho'}Z' + \sqrt{1 - \rho'}\varepsilon'_k \quad (13)$$

with

$$Z' = \phi Z + \sqrt{1 - \phi^2} u, \quad (14)$$

where  $\rho, \rho' \in (0, 1)$ ,  $\phi \in (-1, 1)$ , and  $Z, \varepsilon_1, \dots, \varepsilon_N, \varepsilon'_1, \dots, \varepsilon'_N, u$  are mutually independent standard Gaussians. Consider next the number of  $X_k$  that fall below some threshold  $X_*$ , and the number of  $X'_k$  below  $X'_*$ :

$$A = \sum_{k=1}^N \mathbb{1}_{\{X_k \leq X_*\}}, \quad \text{and} \quad A' = \sum_{k=1}^N \mathbb{1}_{\{X'_k \leq X'_*\}}, \quad (15)$$

where  $X_*$  and  $X'_*$  are constants. Then

$$\text{Cov}(A, A') = N^2 \text{Cov}(p, p'), \quad (16)$$



where

$$p = p(Z) := \mathbb{P}[X_k \leq X_* | Z] = \Phi\left(\frac{X_* - \sqrt{\rho}Z}{\sqrt{1-\rho}}\right), \quad \text{and} \quad p' = \mathbb{P}[X'_k \leq X'_* | Z'] = p'(Z'). \quad (17)$$

Moreover, the correlation between  $A$  and  $A'$  equals the correlation between  $p$  and  $p'$ , in the limit as  $N \rightarrow \infty$ .

*Proof.* We first show that

$$\mathbb{E}[AA'] = \mathbb{E}[\mathbb{E}[A | Z]\mathbb{E}[A' | Z']]. \quad (18)$$

Note that  $A$  is a function of  $Z$  and  $\varepsilon = (\varepsilon_1, \dots, \varepsilon_N)$ , and  $A'$  is a function (indeed, the same function as it happens) of  $Z'$  and  $\varepsilon' = (\varepsilon'_1, \dots, \varepsilon'_N)$ . Now for any non-negative bounded Borel functions  $f$  and  $g$  on  $\mathbb{R}^N$ , and any non-negative bounded Borel functions  $F$  and  $G$  on  $\mathbb{R} \times \mathbb{R}^N$ , we have, on using self-evident notation,

$$\begin{aligned} & \mathbb{E}[f(Z)g(Z')F(Z, \varepsilon)G(Z', \varepsilon')] \\ &= \int f(z)g(\underbrace{\phi z + \sqrt{1-\phi^2}x}_{z'})F(z, y_1, \dots, y_N)G(z', y'_1, \dots, y'_N) d\Phi(z, x, \mathbf{y}, \mathbf{y}') \\ &= \int f(z)g(z') \left[ \left\{ \int F(z, y_1, \dots, y_N) d\Phi(\mathbf{y}) \right\} \left\{ \int G(z', y'_1, \dots, y'_N) d\Phi(\mathbf{y}') \right\} \right] d\Phi(z, x) \\ &= \mathbb{E}[f(Z)g(Z')\mathbb{E}[F(Z, \varepsilon) | Z]\mathbb{E}[G(Z', \varepsilon') | Z']]. \end{aligned} \quad (19)$$

This says that

$$\mathbb{E}[F(Z, \varepsilon)G(Z', \varepsilon') | Z, Z'] = \mathbb{E}[F(Z, \varepsilon) | Z]\mathbb{E}[G(Z', \varepsilon') | Z']. \quad (20)$$

Taking expectation on both sides of equation (20) with respect to  $Z$  and  $Z'$ , we obtain

$$\mathbb{E}[F(Z, \varepsilon)G(Z', \varepsilon')] = \mathbb{E}[\mathbb{E}[F(Z, \varepsilon) | Z]\mathbb{E}[G(Z', \varepsilon') | Z']]. \quad (21)$$

Substituting  $F(Z, \varepsilon) = A$ , and  $G(Z', \varepsilon') = A'$ , we have equation (18) and

$$\begin{aligned} \mathbb{E}[AA'] &= \mathbb{E}[\mathbb{E}[A | Z]\mathbb{E}[A' | Z']] \\ &= \mathbb{E}[NpNp'] = N^2\mathbb{E}[pp'], \end{aligned} \quad (22)$$

The last line is due to the fact that conditional on  $Z$ ,  $A$  is a sum of  $N$  independent indicator variables and follows a binomial distribution with parameters  $N$  and  $\mathbb{E}p$ . Applying (21) again with  $F(Z, \varepsilon) = A$ , and  $G(Z', \varepsilon') = 1$ , or indeed, much more directly by repeated expectations, we have

$$\mathbb{E}[A] = N\mathbb{E}[p], \quad \text{and} \quad \mathbb{E}[A'] = N\mathbb{E}[p']. \quad (23)$$

Hence we conclude that

$$\begin{aligned} \text{Cov}(A, A') &= \mathbb{E}(AA') - \mathbb{E}[A]\mathbb{E}[A'] \\ &= N^2\mathbb{E}[pp'] - N^2\mathbb{E}[p]\mathbb{E}[p'] \\ &= N^2 \text{Cov}(p, p'). \end{aligned}$$

We have

$$\begin{aligned}
\text{Var}(A) &= \mathbb{E} [\mathbb{E}[A^2 | Z]] - N^2(\mathbb{E}[p])^2 \\
&= \mathbb{E}[Np + N(N-1)p^2] - N^2(\mathbb{E}[p])^2 \\
&= N\mathbb{E}[p(1-p)] + N^2 \text{Var}(p).
\end{aligned} \tag{24}$$

Similarly,

$$\text{Var}(A') = N\mathbb{E}[p'(1-p')] + N^2 \text{Var}(p').$$

Putting everything together, we have for the correlations:

$$\begin{aligned}
\text{Corr}(A, A') &= \frac{N^2 \text{Cov}(p, p')}{\sqrt{N\mathbb{E}[p(1-p)] + N^2 \text{Var}(p)} \sqrt{N\mathbb{E}[p'(1-p')] + N^2 \text{Var}(p')}} \\
&= \frac{\text{Corr}(p, p')}{\sqrt{1 + \frac{\mathbb{E}[p(1-p)]}{N \text{Var}(p)}} \sqrt{1 + \frac{\mathbb{E}[p'(1-p')]}{N \text{Var}(p)}}} \\
&= \text{Corr}(p, p') \quad \text{as } N \rightarrow \infty.
\end{aligned} \tag{25}$$

□

**THEOREM 1 (Vintage Correlation in Default Rates).** *Consider a pool of  $N$  mortgages created at each time  $v$ , where  $N$  is fixed. Suppose within each vintage  $v$ , defaults are governed by a Gaussian copula model as in equations (1), (5), (7), and (8) with common risk factor  $Z_v$  being a zero-mean stationary Gaussian process. Assume further that  $\rho_v = \text{Corr}(X_{v,i}, X_{v,j})$ , the correlation parameter for state variables  $X_{v,i}$  of individual mortgages of vintage  $v$ , is positive. Then,  $A_v$  and  $A_{v'}$ , the numbers of defaults observed at time  $T$  within mortgage vintages  $v$  and  $v'$  are correlated if and only if  $\phi_{v,v'} = \text{Corr}(Z_v, Z_{v'}) \neq 0$ , where  $Z_v$  is the common Gaussian risk factor process. Moreover, in the large portfolio limit,  $\text{Corr}(A_v, A_{v'})$  approaches a limiting value determined by  $\phi_{v,v'}$ ,  $\rho_v$ , and  $\rho_{v'}$ .*

*Proof.* Conditional on the common risk factor  $Z_v$ , the number of defaults  $A_v$  is a sum of  $N$  independent indicator variables and follows a binomial distribution. More specifically,

$$\mathbb{P}(A_v = k | Z_v) = \binom{N}{k} p_v^k (1-p_v)^{N-k} \tag{26}$$

where  $p_v$  is the default probability conditional on  $Z_v$ , i.e.,

$$p_v = \mathbb{P}(\tau_{v,i} \leq \tau^* | Z_v) = \mathbb{P}(X_{v,i} \leq X_v^* | Z_v),$$

with

$$X_v^* = \Phi^{-1}(F_v(T)),$$

where  $F_v(T)$  is the probability of default before the time  $T$ . Then

$$\begin{aligned}
p_v &= \mathbb{P}(X_{v,i} \leq X_v^* | Z_v) \\
&= \mathbb{P}\left(\varepsilon_i \leq \frac{X_v^* - \sqrt{\rho_t} Z_v}{\sqrt{1 - \rho_v}}\right) \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\frac{X_v^* - \sqrt{\rho_t} Z_v}{\sqrt{1 - \rho_v}}} \exp\left(-\frac{x^2}{2}\right) dx \\
&= \Phi(Z_v^*),
\end{aligned} \tag{27}$$

where

$$Z_v^* = \frac{X_v^* - \sqrt{\rho_v} Z_v}{\sqrt{1 - \rho_v}}. \tag{28}$$

Similarly,

$$p_{v'} = \Phi(Z_{v'}^*), \tag{29}$$

where

$$Z_{v'}^* = \frac{X_{v'}^* - \sqrt{\rho_{v'}} Z_{v'}}{\sqrt{1 - \rho_{v'}}}. \tag{30}$$

Note that if  $Z_v$  and  $Z_{v'}$  are jointly Gaussian with correlation coefficient  $\phi_{v,v'}$ , we can write

$$Z_v = \phi_{v,v'} Z_{v'} + \sqrt{1 - \phi_{v,v'}^2} u_{v,v'} \quad \text{for } t > j, \tag{31}$$

where  $u_{v,v'}$  are standard Gaussians that are independent of  $Z_{v'}$ . Combining equation (28), (30) and (31), we have

$$Z_v^* = a \phi_{v,v'} Z_{v'}^* + \frac{X_v^* - b \phi_{v,v'} X_{v'}^*}{\sqrt{1 - \rho_v}} - \frac{\sqrt{\rho_v(1 - \phi_{v,v'}^2)}}{\sqrt{1 - \rho_v}} u_{v,v'}, \tag{32}$$

where

$$a = \sqrt{\frac{\rho_v(1 - \rho_{v'})}{\rho_{v'}(1 - \rho_v)}}, \quad b = \sqrt{\frac{\rho_v}{\rho_{v'}}}.$$

Therefore,

$$\begin{aligned}
\text{Cov}(p_v, p_{v'}) &= \text{Cov}(\Phi(Z_v^*), \Phi(Z_{v'}^*)) \\
&= \text{Cov}\left(\Phi\left(a \phi_{v,v'} Z_{v'}^* + \frac{X_v^* - b \phi_{v,v'} X_{v'}^*}{\sqrt{1 - \rho_v}} - \frac{\sqrt{\rho_v(1 - \phi_{v,v'}^2)}}{\sqrt{1 - \rho_v}} u_{v,v'}\right), \Phi(Z_{v'}^*)\right).
\end{aligned} \tag{33}$$

Since  $a > 0$  as  $\rho_v \in (0, 1)$ , we know that the covariance and the correlation between  $p_v$  and  $p_{v'}$  are determined by  $\phi_{v,v'}$ ,  $\rho_v$ , and  $\rho_{v'}$ . They are nonzero if and only if  $\phi_{v,v'} \neq 0$ . Applying Proposition 1, we know that

$$\text{Corr}(A_v, A_{v'}) = \frac{\text{Corr}(p_v, p_{v'})}{\sqrt{1 + \frac{\mathbb{E}[p_v(1-p_v)]}{N \text{Var}(p_{v'})}} \sqrt{1 + \frac{\mathbb{E}[p_{v'}(1-p_{v'})]}{N \text{Var}(p_v)}}} \quad \forall v \neq v'. \tag{34}$$

Therefore,  $A_v$  and  $A_{v'}$  have nonzero correlation as long as  $p_v$  and  $p_{v'}$  do.  $\square$

Equations (33) and (34) provide closed-form expressions for the serial correlation of default rates  $p_v$  of different vintages and the number of defaults  $A_v$ . However, we cannot directly read from equation (33) how the vintage correlation of default rates depends on the serial correlation parameter  $\phi_{v,v'}$  in the common risk factor. The theorem below shows that this dependence is always positive.

**THEOREM 2 (Dependence on Common Risk Factor).** *Under the same settings as in Theorem 1, assume that both the serial correlation  $\phi_{v,v'}$  of the common risk factor and the individual state variable correlation  $\rho_v$  are always positive. Then the number  $A_v$  of defaults in the vintage- $v$  cohort by time  $T$  is positively correlated with the number  $A_{v'}$  in the vintage- $(v')$  cohort. Moreover, this correlation is an increasing function of the serial correlation parameter  $\phi_{v,v'}$  in the common risk factor.*

*Proof.* We will use the notation established in Proposition 1. We can assume that  $v \neq v'$ . Recall that in the Gaussian copula model, name  $i$  in the vintage- $v$  cohort defaults by time  $T$  if the standard Gaussian variable  $X_{v,i}$  falls below a threshold  $X_v^*$ . The unconditional default probability is

$$\mathbb{P}[X_{v,i} \leq X_v^*] = \Phi(X_v^*).$$

For the covariance, we have

$$\begin{aligned} \text{Cov}(A_v, A_{v'}) &= \sum_{k,l=1}^N \text{Cov}(\mathbb{1}_{[X_{v,k} \leq X_v^*]}, \mathbb{1}_{[X_{v',l} \leq X_{v'}^*]}) \\ &= N^2 \text{Cov}(\mathbb{1}_{[X \leq X_v^*]}, \mathbb{1}_{[X' \leq X_{v'}^*]}), \end{aligned} \quad (35)$$

where  $X, X'$  are jointly Gaussian, each standard Gaussian, with mean zero and covariance

$$\mathbb{E}[XX'] = \mathbb{E}[X_{v,k}X_{v',l}],$$

which is the same for all pairs  $k, l$ , since  $v \neq v'$ . This common value of the covariance arises from the covariance between  $Z_v$  and  $Z_{v'}$  along with the covariance between any  $X_{v,k}$  with  $Z_v$ ; it is

$$\text{Cov}(X, X') = \phi_j \sqrt{\rho_v \rho_{v'}}. \quad (36)$$

Now since  $X, X'$  are jointly Gaussian, we can express them in terms of two independent standard Gaussians:

$$\begin{aligned} W_1 &:= X, \\ W_2 &:= \frac{1}{\sqrt{1 - \rho_v \rho_{v'} \phi_{v,v'}^2}} [X' - \phi_{v,v'} \sqrt{\rho_v \rho_{v'}} X]. \end{aligned} \quad (37)$$

We can check readily that these are standard Gaussians with *zero* covariance, and

$$\begin{aligned} X &= W_1, \\ X' &= \phi_{v,v'} \sqrt{\rho_v \rho_{v'}} W_1 + \sqrt{1 - \rho_v \rho_{v'} \phi_{v,v'}^2} W_2. \end{aligned} \quad (38)$$

Let

$$\alpha = \phi_{v,v'} \sqrt{\rho_v \rho_{v'}}.$$

The assumption that  $\rho$  and  $\phi_{v,v'}$  are positive (and, of course, less than 1) implies that

$$0 < \alpha < 1.$$

Note that the covariance between  $p_v$  and  $p_{v'}$  can be expressed as

$$\begin{aligned} \text{Cov}(p_v, p_{v'}) &= \mathbb{E}(p_v p_{v'}) - \mathbb{E}(p_v)\mathbb{E}(p_{v'}) \\ &= \mathbb{E} \left[ \mathbb{E} \left[ \mathbf{1}_{\{X_{v,i} \leq X_v^*\}} \middle| Z_v \right] \mathbb{E} \left[ \mathbf{1}_{\{X_{v',i} \leq X_{v'}^*\}} \middle| Z_{v'} \right] \right] - \mathbb{E}(p_v)\mathbb{E}(p_{v'}) \\ &= \mathbb{E} \left[ \mathbf{1}_{\{X_{v,i} \leq X_v^*\}} \mathbf{1}_{\{X_{v',i} \leq X_{v'}^*\}} \right] - \mathbb{E}(p_v)\mathbb{E}(p_{v'}) \\ &= \mathbb{P} \left[ X_{v,i} \leq X_v^*, X_{v',i} \leq X_{v'}^* \right] - \mathbb{E}(p_v)\mathbb{E}(p_{v'}) \\ &= \mathbb{P} \left[ W_1 \leq X_v^*, \alpha W_1 + \sqrt{1 - \alpha^2} W_2 \leq X_{v'}^* \right] - \mathbb{E}(p_v)\mathbb{E}(p_{v'}) \\ &= \int_{-\infty}^{X_v^*} \Phi \left( \frac{X_{v'}^* - \alpha w_1}{\sqrt{1 - \alpha^2}} \right) \varphi(w_1) dw_1 - \mathbb{E}(p_v)\mathbb{E}(p_{v'}), \end{aligned}$$

where  $\varphi(\cdot)$  is the probability density function of the standard normal distribution. The third equality follows from equation (21). The fifth equality follows from equation (38). The unconditional expectation of  $p_v$  is independent of  $\alpha$ , because

$$\begin{aligned} \mathbb{E}(p_v) &= \mathbb{E}(\mathbb{P}(X_{v,i} \leq X_v^* | Z_v)) \\ &= \mathbb{P}(X_{v,i} \leq X_v^*) \\ &= \Phi(X_v^*). \end{aligned} \tag{39}$$

It follows that

$$\begin{aligned} \frac{\partial}{\partial \alpha} \text{Cov}(p_v, p_{v'}) &= \int_{-\infty}^{X_v^*} \varphi \left( \frac{X_{v'}^* - \alpha w_1}{\sqrt{1 - \alpha^2}} \right) \varphi(w_1) \frac{\partial \left( \frac{X_{v'}^* - \alpha w_1}{\sqrt{1 - \alpha^2}} \right)}{\partial \alpha} dw_1 \\ &= \int_{-\infty}^{X_v^*} \varphi \left( \frac{X_{v'}^* - \alpha w_1}{\sqrt{1 - \alpha^2}} \right) \varphi(w_1) \frac{-w_1 + \alpha X_{v'}^*}{(1 - \alpha^2)^{\frac{3}{2}}} dw_1 \\ &= -\frac{1}{(1 - \alpha^2)^{\frac{3}{2}}} \int_{-\infty}^{X_v^*} (w_1 - \alpha X_{v'}^*) \varphi \left( \frac{X_{v'}^* - \alpha w_1}{\sqrt{1 - \alpha^2}} \right) \varphi(w_1) dw_1. \end{aligned} \tag{40}$$

The last two terms in the integrand can be rewritten as

$$\begin{aligned}
\varphi\left(\frac{X_{v'}^* - \alpha w_1}{\sqrt{1 - \alpha^2}}\right) \varphi(w_1) &= \frac{1}{2\pi} \exp\left[-\frac{(X_{v'}^* - \alpha w_1)^2}{2(1 - \alpha^2)} - \frac{w_1^2}{2}\right] \\
&= \frac{1}{2\pi} \exp\left[-\frac{X_{v'}^{*2} - 2\alpha X_{v'}^* w_1 + w_1^2 \alpha^2 + w_1^2(1 - \alpha^2)}{2(1 - \alpha^2)}\right] \\
&= \frac{1}{2\pi} \exp\left[-\frac{w_1^2 - 2\alpha X_{v'}^* w_1 + \alpha^2 X_{v'}^{*2} + X_{v'}^{*2}(1 - \alpha^2)}{2(1 - \alpha^2)}\right] \\
&= \frac{1}{2\pi} \exp\left[-\frac{(w_1 - \alpha X_{v'}^*)^2 + X_{v'}^{*2}(1 - \alpha^2)}{2(1 - \alpha^2)}\right].
\end{aligned} \tag{41}$$

Substituting equation (41) into (40), we have

$$\frac{\partial}{\partial \alpha} \text{Cov}(p_v, p_{v'}) = -\frac{\exp\left(-\frac{X_{v'}^{*2}}{2}\right)}{2\pi(1 - \alpha^2)^{\frac{3}{2}}} \int_{-\infty}^{X_v^*} (w_1 - \alpha X_{v'}^*) \exp\left[-\frac{(w_1 - \alpha X_{v'}^*)^2}{2(1 - \alpha^2)}\right] dw_1.$$

Make a change of variable and let

$$y := \frac{w_1 - \alpha X_{v'}^*}{\sqrt{1 - \alpha^2}}.$$

It follows that

$$\begin{aligned}
\frac{\partial}{\partial \alpha} \text{Cov}(p_v, p_{v'}) &= -\frac{\exp\left(-\frac{X_{v'}^{*2}}{2}\right)}{2\pi\sqrt{1 - \alpha^2}} \int_{-\infty}^{\frac{X_v^* - \alpha X_{v'}^*}{\sqrt{1 - \alpha^2}}} y \exp\left(-\frac{y^2}{2}\right) dy \\
&= \frac{\exp\left(-\frac{X_{v'}^{*2}}{2}\right)}{2\pi\sqrt{1 - \alpha^2}} \exp\left[-\frac{(X_v^* - \alpha X_{v'}^*)^2}{2(1 - \alpha^2)}\right] \\
&= \frac{1}{2\pi\sqrt{1 - \alpha^2}} \exp\left(-\frac{X_v^{*2} - 2\alpha X_v^* X_{v'}^* + X_{v'}^{*2}}{2(1 - \alpha^2)}\right) > 0.
\end{aligned}$$

Thus, we have shown that the partial derivative of the covariance with respect to  $\alpha$  is positive. Since

$$\alpha = \sqrt{\rho_v \rho_{v'}} \phi_{v, v'},$$

with  $\rho_s$  and  $\phi_{v, v'}$  assumed to be positive, we know that the partial derivatives of the covariance with respect to  $\phi_{v, v'}$ ,  $\rho_v$  and  $\rho_{v'}$  are also positive everywhere. Note that the unconditional variance of  $p_v$  is independent of  $\phi_{v, v'}$  (although dependent of  $\rho_s$ ), which can be seen from equation (27). It follows that the serial correlation of  $p_v$  has positive partial derivative with respect to  $\phi_{v, v'}$ . Recall equation (33), which shows that the covariance of  $p_v$  and  $p_{v'}$  is zero for any value of  $\rho_s$  when  $\phi_{v, v'} = 0$ . This result together with the positive partial derivatives of the covariance with respect to  $\phi_{v, v'}$  ensure that the covariance and thus the vintage correlation of  $p_v$  and  $p_{v'}$  is always positive. From equation (34), noticing the fact that both the expectation and variance of  $p_v$  are independent of  $\phi_{v, v'}$ , we know that

the correlation between  $A_v$  and  $A_{v'}$  must also be positive everywhere and monotonically increasing in  $\phi_{v,v'}$ .  $\square$

So far we have shown that the vintage correlation of  $p_v$  and thus  $A_t$  is positive and increasing in  $\phi_{v,v'}$ . Due to the complexity of the analytical form of the vintage correlation, we resort to numerical methods to study its exact magnitude. We generate plots of the vintage correlation of  $p_v$  as a function of  $\phi_{v,v'}$ . For simplicity, we assume  $\rho_v = \rho_{v'} = \rho$ , and  $X_v^* = X_{v'}^* = X^*$ . Different values of  $X_v^*$  ranging from  $-3$  to  $3$  are implemented. Two of these plots can be seen in Figures 2 and 3. Others are omitted for brevity as they look very similar. In all these plots, vintage correlation is always positive for  $\phi_{v,v'} \in (0, 1)$  and monotonically increasing in  $\phi_{v,v'}$ , matching our theoretical findings. Moreover, the magnitude of vintage correlation is always close to  $\phi_{v,v'}$ , although when the absolute value of  $X_v^*$  increases, the function becomes more convex.

#### 4. MONTE CARLO SIMULATIONS

In this section, we study the link between serial correlation in a common risk factor and vintage correlation in pools of mortgages in two sets of simulations: First, a series of mortgage pools is simulated to confirm the analytical results of Section 3. Second, a waterfall structure is simulated to study temporal correlation in MBS.

**4.1. Vintage Correlation in Mortgage Pools.** We conduct a Monte Carlo simulation to study how serial correlation of a common risk factor propagates into vintage correlation in default rates. We simulate default times for individual mortgages according to equations (1), (5), (7), and (8). From the simulated default times, the default rate of a pool of mortgages is calculated.

In each simulation, we construct a cohort of  $N = 100$  homogeneous mortgages in every month  $v = 1, 2, \dots, 120$ . We simulate a monthly time series of the common risk factor  $Z_v$ , which is assumed to have an AR(1) structure with unconditional mean zero and variance one,

$$Z_v = \phi Z_{v-1} + \sqrt{1 - \phi^2} u_v \quad \forall v = 2, 3, \dots, 120. \quad (42)$$

The errors  $u_v$  are i.i.d. standard Gaussian. The initial observation  $Z_1$  is a standard normal random variable. We report the case where  $\phi = 0.95$ . We choose  $\phi$  close to one due to the observation that the autocorrelation of the housing index is high.<sup>4</sup> Each mortgage  $i$  issued at time  $v$  has a state variable  $X_{v,i}$  assigned to it that determines its default time. The time series properties of  $X_{v,i}$  follow equation (8). The error  $\varepsilon_i$  in equation (8) is independent of  $u_v$ .

To simulate the actual default rates of mortgages, we need to specify the marginal distribution functions of default times  $F(\cdot)$  as in equation (1). We define a function  $\mathcal{F}(\cdot)$ , which takes a time period as argument and returns the default probability of a mortgage within that time period since its initiation. We assume that this  $\mathcal{F}(\cdot)$  is fixed across different vintages, which means that mortgages of different cohorts have a same unconditional default probability in the next  $S$  periods from their initiation, where  $S = 1, 2, \dots$ . It is easy to verify that  $F_v(T) = \mathcal{F}(T - v)$ . The values of the function  $\mathcal{F}(\cdot)$  are specified

<sup>4</sup>Other values of  $\phi$  are also tried, but not reported here. The results are all consistent with our theoretical findings.

in Table 1, for both subprime and prime mortgages. Intermediate values of  $\mathcal{F}(\cdot)$  are linearly interpolated from this table. While these values are in the same range as actual default rates of subprime and prime mortgages in the last ten years, their specification is rather arbitrary as it has little impact on the stochastic structure of the simulated default rates. We set the observation time  $T$  to be 144, which is two years after the creation of the last vintage, as we need to give the last vintage some time window to have possible default events. For example, in each month from January 1998 to December 2007, 100 mortgages are created. Then in December 2009, we examine the default rates of these mortgages within each vintage.

We need to consider two cases, subprime and prime. For the subprime case, every vintage is given a two-year window to default, so the unconditional default probability is constant across vintages. On the other hand, prime mortgages have decreasing default probability through subsequent vintages. For example, in our simulation, the first vintage has a time window of 144 months to default, the second vintage has 143 months, the third has 142 months, and so on. Therefore, older vintages are more likely to default by observation time  $T$  than newer vintages. This is why the fixed ex-post observation time of defaults is one difference that distinguishes vintage correlation from serial correlation.

We construct a time series  $\tau_{v,i}$  of default times of mortgage  $i$  issued at time  $v$  according to equation (5).<sup>5</sup> Time series of default rates  $\bar{A}_v$  are computed as:

$$\bar{A}_v(\tau_v^*) = \frac{\#\{\text{mortgages for which } \tau_{v,i} \leq \tau_v^*\}}{N}.$$

In the subprime case,  $\tau_v^* = 24$  is set to be constant. In the prime case,  $\tau_v^* = T - v$  varies across vintages.

The simulation is repeated 1000 times. For the subprime case, the average simulated default rates are plotted in Figure 4. For the prime case, average simulated default rates are plotted in Figure 5. Note that because of the decreasing time window to default, the default rates in Figure 5 have a decreasing trend.

In the subprime case, we can use the sample autocorrelation and partial autocorrelation functions to estimate vintage correlation, because the unconditional default probability is constant across vintages, so that averaging over different vintages and averaging over different pools is the same. In the prime case, we have to calculate vintage correlation proper. Since we have 1000 Monte Carlo observations of default rates for each vintage, we can calculate the correlation between two vintages using those samples. For the partial autocorrelation function, we simply demean the series of default rates and obtain the usual partial autocorrelation function.

We plot the estimated vintage correlation in the second rows of Figure 4 and 5 for subprime and prime cases, respectively. As can be seen, the correlation of the default rates of the first vintage with older vintages decreases geometrically. In both cases, the estimated first-order coefficient of default rates is close to but less than  $\phi = 0.95$ , the AR(1) coefficient of the common risk factor. The partial

<sup>5</sup>Note that this is not a time series of default times for a single mortgage, since a single mortgage defaults only once or never. Rather, the index  $i$  is a placeholder for a position in a mortgage pool. In this sense,  $\tau_{v,i}$  is the time series of default times of mortgages in position  $i$  in the pool over vintages  $v$ .



autocorrelation functions are plotted in the third rows of Figures 4 and 5. They are significant only at lag one. This phenomenon is also observed when we set  $\phi$  to other values. Both the sample autocorrelation and partial autocorrelation functions indicate that the default rates follow a first-order autoregressive process, similar to the specification of the common risk factor. However, compared with the subprime case, the default rates of prime mortgages seem to have longer memory.

The similarity between the magnitude of the autocorrelation coefficient of default rates and common risk factor can be explained by the following Taylor expansion. Taylor-expanding equation (27) at  $Z_v^* = 0$  to first order, we have

$$p_v \approx \frac{1}{2} + \frac{1}{\sqrt{2\pi}} Z_v^* \quad (43)$$

Since  $p_v$  is in this sense approximately a linear transformation of  $Z_v^*$ , which is a linear transformation of  $Z_t$ , it approximately follows a stochastic process that has the same serial correlation as  $Z_t$ .

**4.2. Vintage Correlation in Waterfall Structures.** Using the Gaussian copula approach, we have already shown that the time series of default rates in mortgage pools inherits vintage correlation from the serial correlation of the common risk factor. We now study how this affects the performance of assets such as MBS that are securitized from the mortgage pool in a so-called waterfall.

The basic elements of the simulation are:

- (1) a time line of 120 months and an observation time  $T = 144$ ;
- (2) a mortgage contract with a principal of \$1 and a maturity of 15 years. The annual interest rate on the mortgage loan is 9%. Fixed monthly payments are received until the mortgage defaults or is paid in full. In each month, a pool of 100 such mortgages is created.
- (3) A pool of 100 units of MBS is securitized from the mortgage cohort in each month. Every unit of MBS has a principal of \$1. There are four tranches in our structure: the senior tranche, the mezzanine tranche, the subordinate tranche, and the equity tranche. The senior tranche consists of the top 70% of the face value of all mortgages created in each month (that is, there are 70 units of senior MBS); the mezzanine tranche consists of the next 25%; the subordinate tranche consist of the next 4%; the equity tranche has the bottom 1%. Each senior MBS pays an annual interest rate of 6%; each mezzanine MBS pays 15%; each subordinate MBS pays 20%. The equity tranche does not pay interest but retains residual profits, if any.

The basic setup of the simulation is illustrated in Figure 6. For a cohort of mortgages issued at time  $v$  and the MBS derived from it, the securitization process works as follows. At the end of each month, each mortgage either defaults or makes a fixed monthly payment. The method to determine default is the same that we have used before: mortgage  $i$  issued at time  $v$  defaults at  $\tau_{v,i}$ , which is generated by the Gaussian copula approach according to equations (1), (5), (7), and (8). We consider both subprime and prime scenarios, as in the case of default rates. For subprime mortgages, we assume that each individual mortgage receives a prepayment of the outstanding principal at the end of the teaser period if it has not defaulted, so the default events and cash flows only happen within the teaser period. For the prime case, there is no such restriction. Again, we assume the common risk factor to follow an AR(1)

process with first-order autocorrelation coefficient  $\phi = 0.95$ . The cross-name correlation coefficient  $\rho$  is set to be 0.5. The unconditional default probabilities over time are obtained from Table 1.

If a mortgage has not defaulted, the interest payments received from it are used to pay the interest specified on the MBS from top to bottom. Thus, the cash inflow is used to pay the senior tranche first (6% of the remaining principal of the senior tranche at the beginning of the month). The residual amount, if any, is used to pay the mezzanine tranche, after that the subordinate tranche, and any still remaining funds are collected in the equity tranche. If the cash inflow passes a tranche threshold but does not cover the following tranche, it is prorated to the following tranche. Any residual funds after all the non-equity tranches have been paid add to the principal of the equity tranche. Principal payments are processed analogously. We assume a recovery rate of 50% on the outstanding principal for defaulted mortgages. The 50% loss of principal is deducted from the principal of the lowest ranked outstanding MBS. This means that the equity tranche covers the principal loss first. If there is no principal left in the equity tranche, the subordinate tranche covers the remaining loss and so on upwards. In order to capture MBS price performance through time, we calculate the present value of all cash flows received from each MBS tranche and average across 1000 simulations. This results in a time series of expected present values and can be viewed as a proxy for tranche price evolution.

Before we examine the vintage correlation of the present value of MBS tranches, we look at the time series of total principal loss across MBS tranches. In our simulations, no loss of principal occurred for the senior tranche. The series of expected principal losses of other tranches and their sample autocorrelation and sample partial autocorrelation are plotted in Figures 7 and 8 for subprime and prime scenarios respectively. We use the same method to obtain the autocorrelation functions for prime mortgages as in the case of default rates. The correlograms show that the expected loss of principal for each tranche follows an AR(1) process, although the estimated coefficients are smaller than  $\phi = 0.95$ , the first-order autocorrelation coefficient of  $Z_v$ , in all cases.

The series of present values of cash flows for each tranche and their sample autocorrelation and partial autocorrelation functions are plotted in Figures 9 and 10 for subprime and prime scenarios, respectively. The senior tranche displays a significant first-order autocorrelation coefficient due to losses in interest payments although there are no losses in principal. The partial autocorrelation functions, which have significant positive values for more than one lag, suggest that the cash flows may not follow an AR(1) process due to the high non-linearity. However, the estimated vintage correlation still decreases over vintages, same as in an AR(1) process, which indicates that our findings for default rates can be extended to cash flows.

## 5. CONCLUSIONS

Default rates of subprime mortgages exhibit temporal correlation. Default events of subprime-mortgages depend on house price changes that are serially correlated, and this serial correlation is inherited by the sequence of defaults. We model a house price index by a common risk factor in a

Gaussian copula model. Analytical findings and simulations show that credit assets inherit correlation of defaults across different vintages of initiation from serial correlation in the common risk factor. The same is true in waterfall structures, used in mortgage-backed securities and collateralized debt obligations. Simulation results demonstrate that when the common risk factor of different cohorts of individual mortgages is serially correlated, the price of the MBS securitized from these mortgages also displays serial correlation. These findings are consistent with the view that the current financial crisis, which was triggered by a large and serially correlated arrival of subprime mortgage defaults, has its origin in the decline of the housing market.

#### ACKNOWLEDGMENTS

We thank seminar participants at Stanford University and Darrell Duffie for helpful discussions.

#### REFERENCES

- ANDERSEN, L., AND J. SIDENIUS (2005): "Extensions to the Gaussian Copula: Random Recovery and Random Factor Loadings," *Journal of Credit Risk*, 1(1), 29–70.
- ANDERSEN, L., J. SIDENIUS, AND S. BASU (2003): "All your Hedges in One Basket," *Risk*, November, 67–72.
- BAJARI, P., C. CHU, AND M. PARK (2008): "An Empirical Model of Subprime Mortgage Default From 2000 to 2007," Working Paper 14358, NBER.
- BEARE, B. (2010): "Copulas and Temporal Dependence," *Econometrica*, 78(1), 395–410.
- BERD, A., R. ENGLE, AND A. VORONOV (2007): "The Underlying Dynamics of Credit Correlations," *Journal of Credit Risk*, 3(2), 27–62.
- BLUHM, C., L. OVERBECK, AND C. WAGNER (2002): *An Introduction to Credit Risk Modeling*, Financial Mathematics Series. Chapman & Hall/CRC, London, UK.
- BURTSCHHELL, X., J. GREGORY, AND J.-P. LAURENT (2009): "A Comparative Analysis of CDO Pricing Models," *Journal of Derivatives*, 16(4), 9–37.
- CABALLERO, R. J., AND A. KRISHNAMURTHY (2009): "Global Imbalances and Financial Fragility," *American Economic Review*, 99(2), 584–588.
- CROUHY, M. G., R. A. JARROW, AND S. M. TURNBULL (2008): "The Subprime Credit Crisis of 2007," *Journal of Derivatives*, 16(1), 81–110.
- DAGLISH, T. (2009): "What Motivates a Subprime Borrower to Default?," *Journal of Banking and Finance*, 33(4), 681–693.
- DAS, S. R., L. FREED, G. GENG, AND N. KAPADIA (2006): "Correlated Default Risk," *Journal of Fixed Income*, 16(2), 7–32.
- DUFFIE, D., AND K. SINGLETON (2003): *Credit Risk: Pricing, Measurement, and Management*, Princeton Series in Finance. Princeton University Press, Princeton, NJ.
- FIGLEWSKI, S. (2009): "Viewing the Crisis from 20,000 Feet Up," *Journal of Derivatives*, 16(3), 53–61.
- FREES, E. W., AND E. A. VALDEZ (1998): "Understanding Relationships Using Copulas," *North American Actuarial Journal*, 2(1), 1–25.
- FREY, R., AND A. MCNEIL (2003): "Dependent Defaults in Models of Portfolio Credit Risk," *Journal of Risk*, 6(1), 59–92.
- GIESECKE, K. (2003): "A Simple Exponential Model for Dependent Defaults," *Journal of Fixed Income*, 13(3), 74–83.
- GORTON, G. (2008): "The Panic of 2007," Working Paper 14358, NBER.
- (2009): "Information, Liquidity, and the (Ongoing) Panic of 2007," *American Economic Review*, 99(2), 567–572.
- GRAZIANO, G. D., AND L. C. G. ROGERS (2009): "A Dynamic Approach to the Modeling of Correlation Credit Derivatives Using Markov Chains," *International Journal of Theoretical and Applied Finance*, 12(1), 45–62.

- HAYRE, L. S., M. SARAF, R. YOUNG, AND J. CHEN (2008): "Modeling of Mortgage Defaults," *Journal of Fixed Income*, 17(4), 6–30.
- HULL, J., M. PREDESCU, AND A. WHITE (2009): "The Valuation of Correlation-Dependent Credit Derivatives Using a Structural Model," Working paper.
- HULL, J., AND A. WHITE (2001): "Valuing Credit Default Swaps II: Modeling Default Correlations," *Journal of Derivatives*, 8(3), 12–22.
- LANDO, D. (2004): *Credit Risk Modeling: Theory and Applications*, Princeton Series in Finance. Princeton University Press, Princeton, NJ.
- LAURENT, J., AND J. GREGORY (2005): "Basket Default Swaps, CDOs and Factor Copulas," *Journal of Risk*, 7(4), 103–122.
- LI, D. (2000): "On Default Correlation: A Copula Function Approach," *Journal of Fixed Income*, 9(4), 43–54.
- LINDSKOG, F., AND A. MCNEIL (2003): "Common Poisson Shock Models: Applications to Insurance and Credit Risk Modelling," *ASTIN Bulletin*, 33(2), 209–238.
- MURPHY, D. (2008): "A Preliminary Enquiry into the Causes of the Credit Crunch," *Quantitative Finance*, 8(5), 435–451.
- REINHART, C. M., AND K. S. ROGOFF (2008): "Is the US Sub-Prime Financial Crisis So Different? An International Historical Comparison," *American Economic Review*, 98(2), 339–344.
- (2009): "The Aftermath of Financial Crises," *American Economic Review*, 99(2), 466–472.
- SCHÖNBUCHER, P. (2003): *Credit Derivatives Pricing Models: Models, Pricing and Implementation*. Wiley, Chichester, UK.
- SCHÖNBUCHER, P., AND D. SCHUBERT (2001): "Copula Dependent Default Risk in Intensity Models," Working paper, Bonn University.
- SERVIGNY, A., AND O. RENAULT (2002): "Default Correlation: Empirical Evidence," Working paper, Standard and Poor's.
- SIDENIUS, J., V. PITERBARG, AND L. ANDERSEN (2008): "A New Framework for Dynamic Credit Portfolio Loss Modelling," *International Journal of Theoretical and Applied Finance*, 11(2), 163–197.
- SKLAR, A. (1959): "Fonctions de Repartition a  $n$  Dimensions et Leurs Marges," *Publ. Inst. Stat. Univ. Paris*, 8, 229–231.

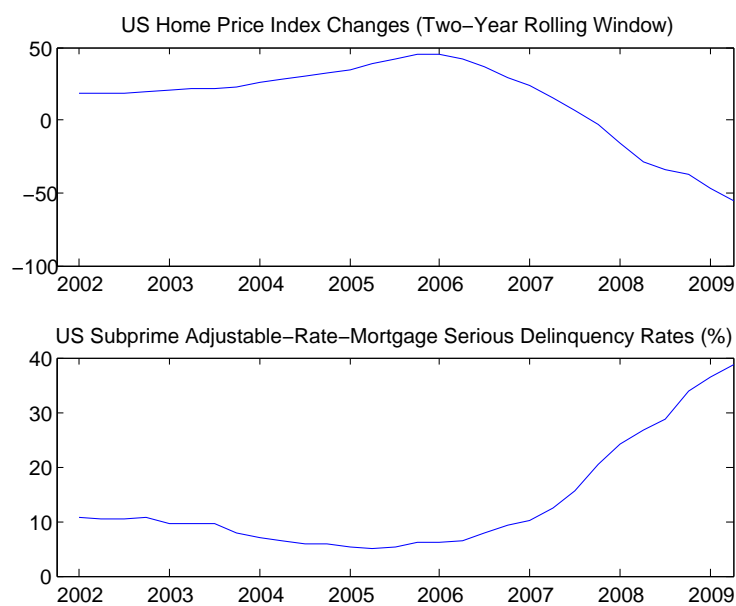
TABLE 1. Default Probabilities Through Time ( $F(\tau)$ ).

Subprime					
Time (Month)	12	24	36	72	144
Default Probability	0.04	0.10	0.12	0.13	0.14

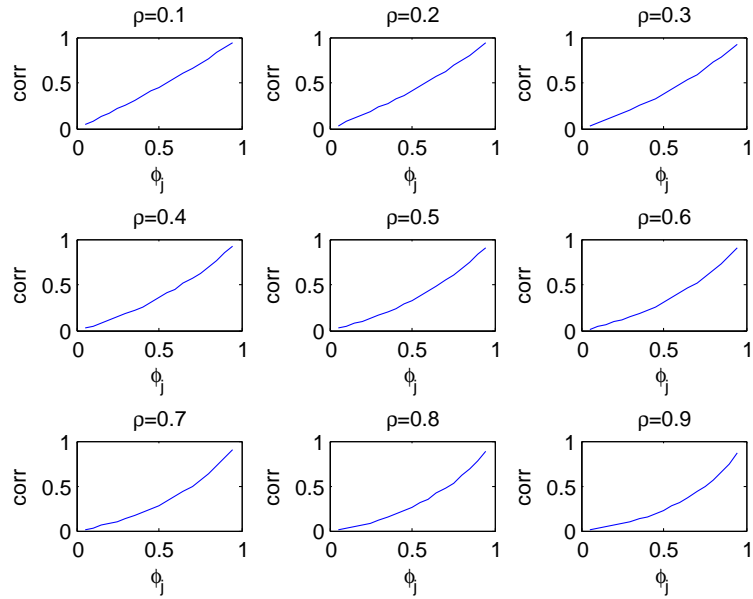
  

Prime					
Time (Month)	12	24	36	72	144
Default Probability	0.01	0.02	0.03	0.04	0.05

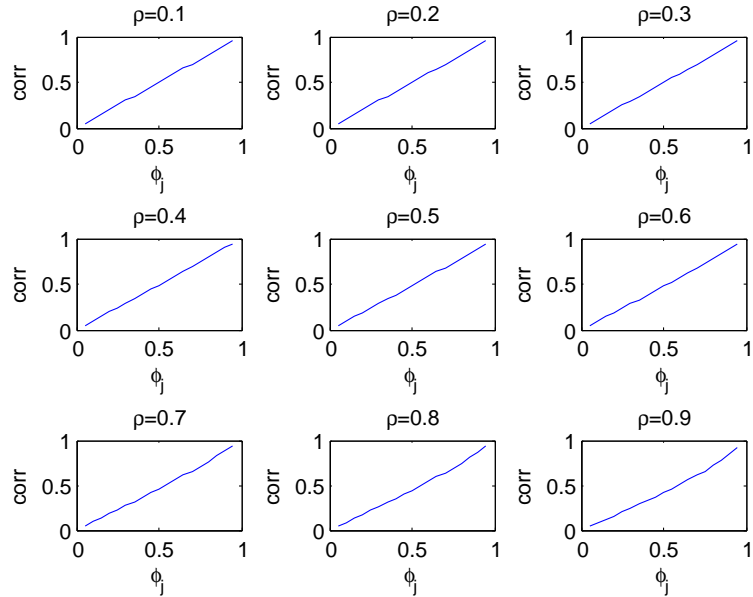
FIGURE 1. Two-Year Changes in U.S. House Price and Subprime ARM Serious Delinquency Rates



“U.S. home price two-year rolling changes” are two-year overlapping changes in the S&P Case-Shiller U.S. National Home Price index. “Subprime ARM Serious Delinquency Rates” are obtained from the Mortgage Banker Association. Both series cover the first quarter in 2002 to the second quarter in 2009.

FIGURE 2. **Correlation of  $p_v$  and  $p_{v'}$  when  $X^* = -2$** 

Correlation between  $p_v$  and  $p_{v'}$  as a function of  $\phi$  for different scenarios of the cross-mortgage correlation parameter  $\rho \in (0, 1)$ .  $X^*$  is set to  $-2$  in all scenarios.

FIGURE 3. **Correlation of  $p_v$  and  $p_{v'}$  when  $X^* = 0$** 

Correlation between  $p_v$  and  $p_{v'}$  as a function of  $\phi$  for different scenarios of the cross-mortgage correlation parameter  $\rho \in (0, 1)$ .  $X^*$  is set to 0 in all scenarios.

FIGURE 4. **Serial Correlation in Default Rates of Subprime Mortgages**

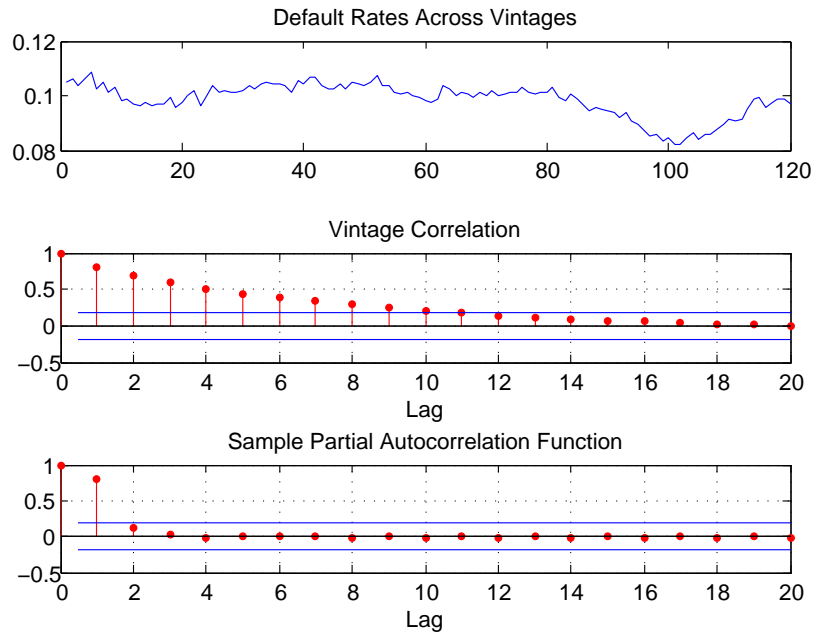


FIGURE 5. **Serial Correlation in Default Rates of Prime Mortgages**

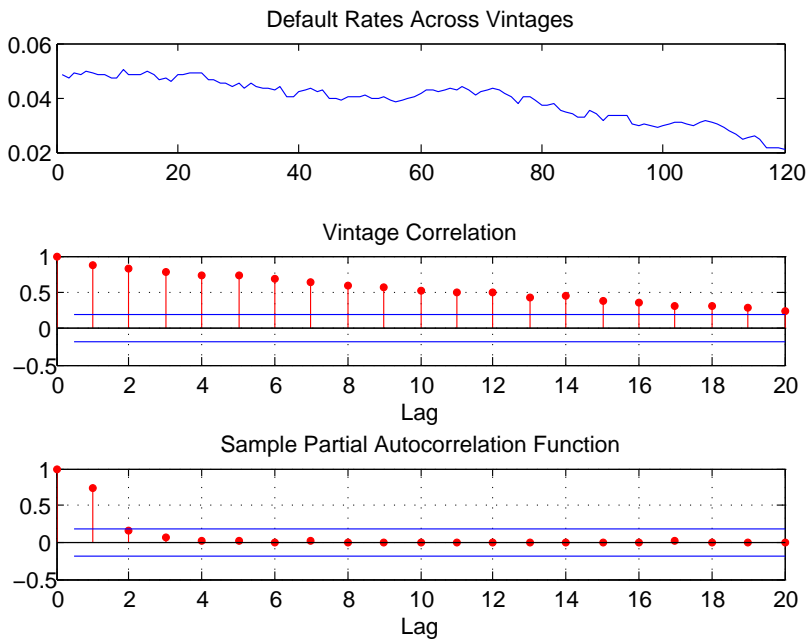


FIGURE 6. Simulated Mortgages and MBS

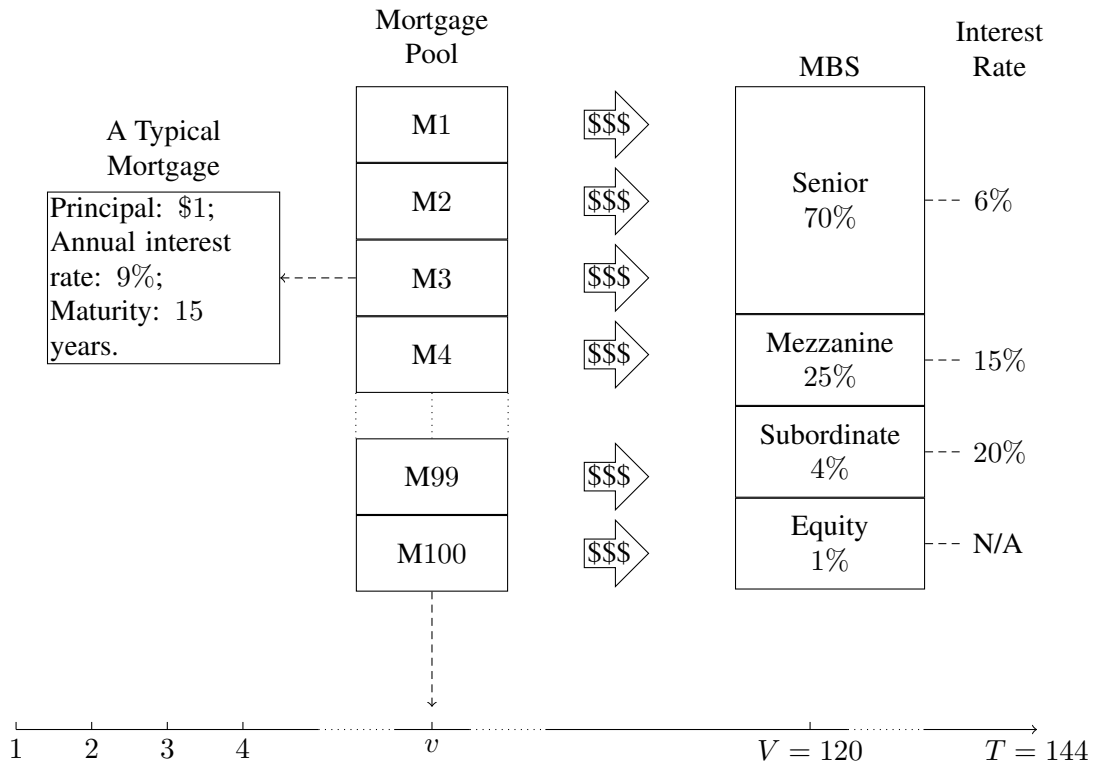
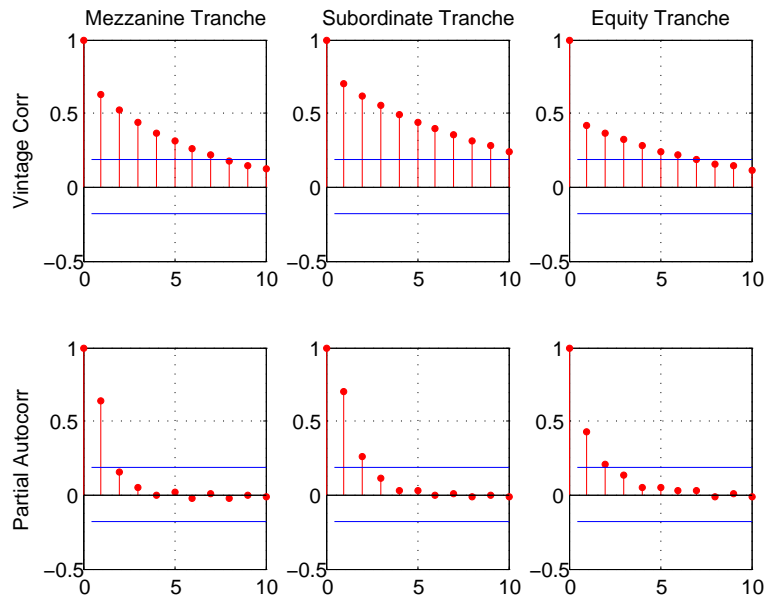


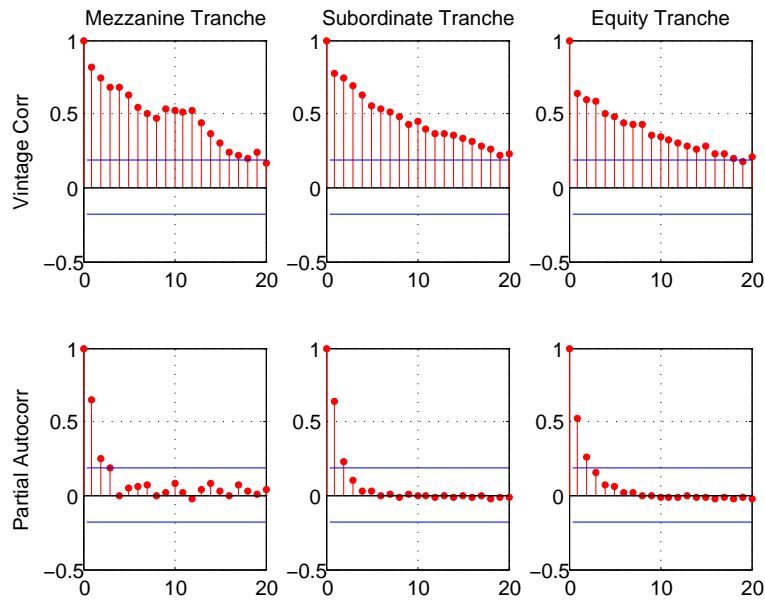


FIGURE 7. **Serial Correlation in Principal Losses of Subprime MBS**



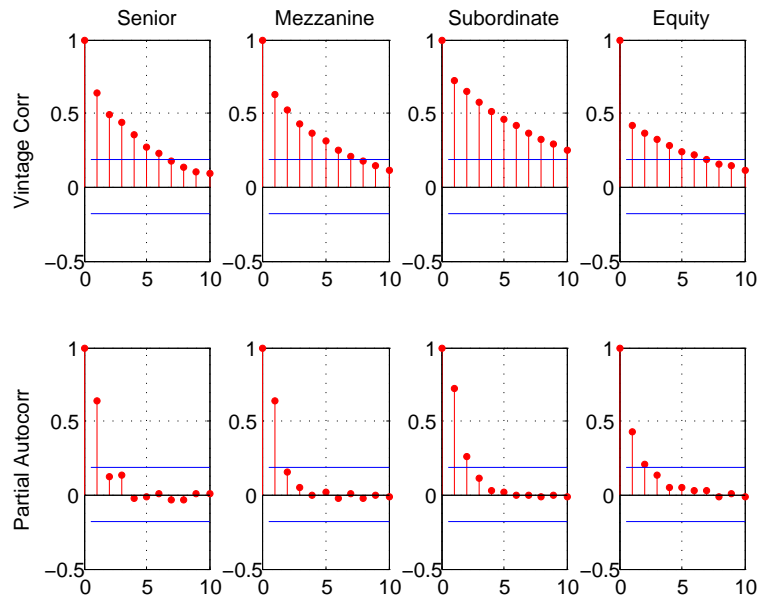
The first row plots the vintage correlation of the principal loss of each tranche. The correlation is estimated using the sample autocorrelation function. The second row plots the partial autocorrelation functions.

FIGURE 8. **Serial Correlation in Principal Losses of Prime MBS**



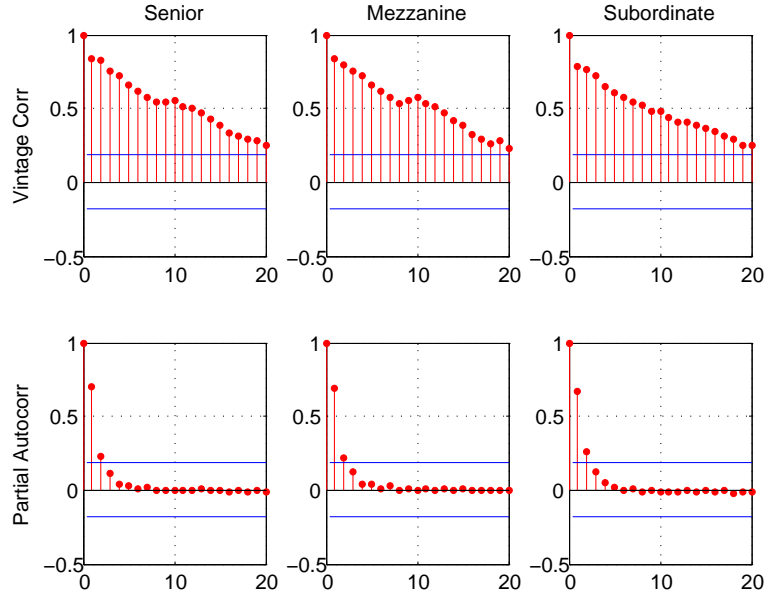
The first row plots the vintage correlation of the principal loss of each tranche. The correlation is estimated using the correlation between the first and subsequent vintages, each of which has a Monte Carlo sample size of 1000. The second row plots the partial autocorrelation functions of the demeaned series of principal losses.

FIGURE 9. Cash Flows from Subprime MBS



The first row plots the vintage correlation of the cash flow received by each tranche. The correlation is estimated using the sample autocorrelation function. The second rows plot the partial autocorrelation functions.

FIGURE 10. Cash Flows from Prime MBS



The first row plots the vintage correlation of the cash flow received by each tranche. The correlation is estimated using the correlation between the first and subsequent vintages, each of which has a Monte Carlo sample size of 1000. The second row plots the partial autocorrelation functions of the demeaned series of cash flow.

(E. Hillebrand) DEPARTMENT OF ECONOMICS, LOUISIANA STATE UNIVERSITY, BATON ROUGE, USA.  
*E-mail address:* erhil@lsu.edu

(A. Sengupta) DEPARTMENT OF MATHEMATICS, LOUISIANA STATE UNIVERSITY, BATON ROUGE, USA.  
*E-mail address:* ambarnsg@yahoo.com

(J. Xu) DEPARTMENT OF ECONOMICS, LOUISIANA STATE UNIVERSITY, BATON ROUGE, USA.  
*E-mail address:* jxu2@tigers.lsu.edu