Communications on Stochastic Analysis

Volume 6 | Number 3 Article 9

9-1-2012

Temporal correlation of defaults in subprime securitization

Eric Hillebrand

Ambar N Sengupta

Junyue Xu

Communications on Stochastic Analysis
Vol. 6, No. 3 (2012) 487-511

Serials Publications
www.serialspublications.com

TEMPORAL CORRELATION OF DEFAULTS IN SUBPRIME SECURITIZATION

ERIC HILLEBRAND, AMBAR N. SENGUPTA, AND JUNYUE XU

ABSTRACT. We examine the subprime market beginning with a subprime mortgage, followed by a portfolio of such mortgages and then a series of such portfolios. We obtain an explicit formula for the relationship between loss distribution and seniority-based interest rates. 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 using a waterfall structure and find that subsequent vintages of these securities inherit temporal correlation from the common risk factor.

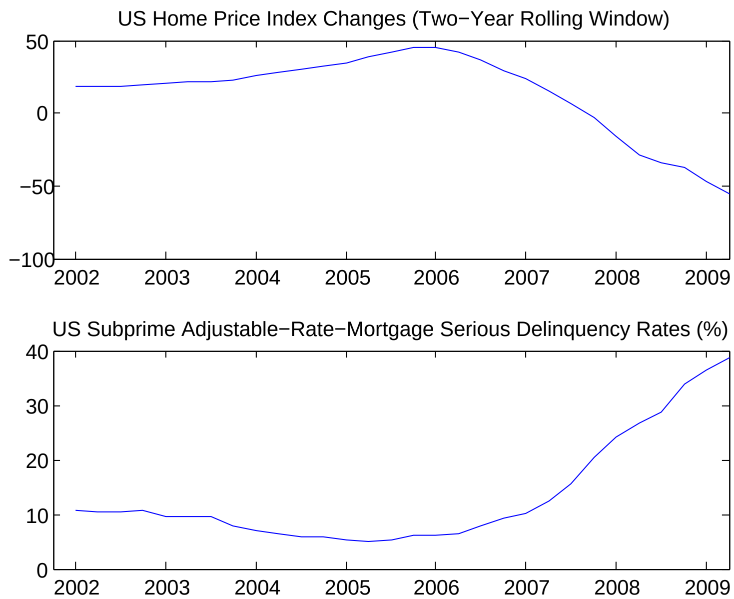
1. Introduction

In this paper we (i) derive closed-form mathematical formulas (4.3) and (4.12) connecting interest rates paid by tranches of Collateralized Debt Obligations (CDOs) and the corresponding loss distributions, (ii) present a two-step Gaussian copula model (Proposition 6.1) governing correlated CDOs, and (iii) study the behavior of correlated CDOs both mathematically and through simulations. The context and motivation for this study is the investigation of mortgage backed securitized structures built out of subprime mortgages that were at the center of the crisis that began in 2007. Our investigation demonstrates, both theoretically and numerically, how the serial correlation in the evolution of the common factor, reflecting the general level of home prices, propagates into a correlated accumulation of losses in tranches of securitized structures based on subprime mortgages of specific vintages. The key feature of these mortgages is the short time horizon to default/prepayment that makes it possible to model the corresponding residential mortgage backed securities (RMBS) as forming one-period CDOs. We explain the difference in behavior between RMBS based on subprime mortgages and those based on prime mortgages in Table 1 and related discussions.

During the subprime crisis, beginning in 2007, subprime mortgages created at different times have defaulted one after another. Figure 1, lower panel, shows the time series of serious delinquency rates of subprime mortgages from 2002 to 2009. (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.) Defaults of subprime mortgages are closely connected to house price fluctuations, as suggested, among others, by [26] (see also [4, 16, 29].) Most subprime mortgages

Received 2012-5-22; Communicated by the editors.
2000 *Mathematics Subject Classification.* 62M10; 91G40.
Key words and phrases. Mortgage-backed securities, CDO, vintage correlation Gaussian copula .

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.

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 [26] 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 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.

We analyze how the dynamics of housing prices propagate, through the dynamics of defaults, to the dynamics of tranche losses in securitized structures based on 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.

One of our objectives in this paper is to provide a formal examination of one of the important causes of the current crisis. (For different perspectives on the causes and effects of the subprime crisis, see also [12, 15, 20, 27, 39, 42, 43].) Vintage correlation in default rates and MBS prices also has implications for asset pricing. To price some derivatives, for example forward starting 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.

2. The Subprime Structure

In a typical subprime mortgage, the loan is amortized over a long period, usually 30 years, but at the end of the first two (or three) years the interest rate is reset to a significantly higher level; a substantial prepayment fee is charged at this time if the loan is paid off. The aim is to force the borrower to repay the loan (and obtain a new one), and the prepayment fee essentially represents extraction of equity from the property, assuming the property has increased in value. If there is sufficient appreciation in the price of the property then both lender and borrower win. However, if the property value decreases then the borrower is likely to default.

Let us make a simple and idealized model of the subprime mortgage cashflow. Let $P_0 = 1$ be the price of the property at time 0, when a loan of the same amount is taken to purchase the property (or against the property as collateral). At time T the price of the property is P_T, and the loan is terminated, resulting either in a prepayment fee k plus outstanding loan amount or default, in which case the lender recovers an amount R. For simplicity of analysis at this stage we assume 0 interest rate up to time T; we can view the interest payments as being built into k or R, ignoring, as a first approximation, defaults prior to time T (for more on early defaults see [7]). The borrower refinances if P_T is above a threshold P_* (say, the present value of future payments on a new loan) and defaults otherwise. Thus the net cashflow to the lender is

$$k\mathbb{1}_{[P_T > P_*]} - (1 - R)\mathbb{1}_{[P_T \leq P_*]}, \tag{2.1}$$

with all payments and values normalized to time-0 money. The expected earning is therefore

$$(k + 1 - R)\mathbb{P}(P_T > P_*) - (1 - R),$$

for the probability measure $\mathbb{P}$ being used. We will not need this expected value but observe simply that a default occurs when $P_T < P_*$, and so, if $\log P_T$ is Gaussian then default occurs for a particular mortgage if a suitable standard Gaussian variable takes a value below some threshold.

It is clear that nothing like the above model would apply to prime mortgages. The main risk (for the lender) associated to a long-term prime mortgage is that of prepayment, though, of course, default risk is also present. A random prepayment time embedded into the amortization schedule makes it a different problem to value a prime mortgage. In contrast, for the subprime mortgage the lender is relying on the prepayment fee and even the borrower hopes to extract equity on the property through refinancing under the assumption that the property value goes up in the time span $[0, T]$. (The prepayment fee feature has been controversial; see, for example, [14, page 50-51].) We refer to the studies [7, 14, 26] for details on the economic background, evolution and ramifications of the subprime mortgage market, which went through a major expansion in the mid 1990s.

3. Portfolio Default Model

Securitization makes it possible to have a much larger pool of potential investors in a given market. For mortgages the securitization structure has two sides: (i) assets are mortgages; (ii) liabilities are debts tranched into seniority levels. In this section we briefly examine the default behavior in a portfolio of subprime mortgages (or any assets that have default risk at the end of a given time period). In Section 4 we will examine a model structure for distributing the losses resulting from defaults across tranches.

For our purposes consider N subprime mortgages, issued at time 0 and (p)repaid or defaulting at time T, each of amount 1. In this section, for the sake of a qualitative understanding we assume 0 recovery, and that a default translates into a loss of 1 unit (we neglect interest rates, which can be built in for a more quantitatively accurate analysis).

Current models of home price indices go back to the work of Wyngarden [49] , where indices were constructed by using prices from repeated sales of the same property at different times (from which property price changes were calculated). Bailey et al. [3] examined repeated sales data and developed a regression-based method for constructing an index of home prices. This was further refined by Case and Shiller [13] into a form that, in extensions and reformulations, has become an industry-wide standard. The method in [13] is based on the following model for the price P_{iT} of house i at time T:

$$\log P_{iT} = C_T + H_{iT} + N_{iT}, \tag{3.1}$$

where C_T is the log-price at time T across a region (city, in their formulation), H_{iT} is a mean-zero Gaussian random walk (with variance same for all i), and N_{iT} is a house-specific random error of zero mean and constant variance (not dependent on i). The three terms on the right in equation (3.1) are independent and (N_{iT}) is a sale-specific fluctuation that is serially uncorrelated; a variety of correlation structures could be introduced in modifications of the model. We will return to this later in equation (5.7) (with slightly different notation) where we will consider different values of T. For now we focus on a portfolio of N subprime mortgages $i \in \{1, \ldots, N\}$ with a fixed value of T.

Let X_i be the random variable

$$X_i = \frac{\log P_{iT} - m_i}{s_i}, \tag{3.2}$$

where m_i is the mean and s_i is the standard devation of $\log P_{iT}$ with respect to some probability measure of interest (for example, the market pricing risk-neutral measure).

Keeping in mind (3.1) we assume that

$$X_i = \sqrt{\rho} Z + \sqrt{1-\rho}\,\epsilon_i \tag{3.3}$$

for some $\rho > 0$, where $(Z, \epsilon_1, \ldots, \epsilon_N)$ is a standard Gaussian in $\mathbb{R}^{N+1}$, with independent components. Mortgage i defaults when X_i crosses below a threshold X_*, so that the assumed common default probability for the mortgages is

$$\mathbb{P}[X_i < X_*] = \mathbb{E}\left[\mathbb{1}_{[X_i < X_*]}\right]. \tag{3.4}$$

The total number of defaults, or portfolio loss (with our assumptions), is

$$L = \sum_{j=1}^{N} \mathbb{1}_{[X_j < X_*]}. \tag{3.5}$$

The cash inflow at time T is the random variable

$$S(T) = \sum_{j=1}^{N} \mathbb{1}_{[X_j \geq X_*]}. \tag{3.6}$$

Pooling of investment funds and lending them for property mortgages is natural and has long been in practice (see Bogue [9, page 73]). In the modern era Ginnie Mae issued the first MBS in 1970 in "pass through" form which did not protect against prepayment risk. In 1983 Freddie Mac issued Collateralized Mortgage Obligations (CMOs) that had a waterfall-like structure and seniority classes with different maturities. The literature on securitization is vast (see, for instance, [19, 36, 41]).

4. Tranche Securitization: Loss Distribution and Tranche Rates

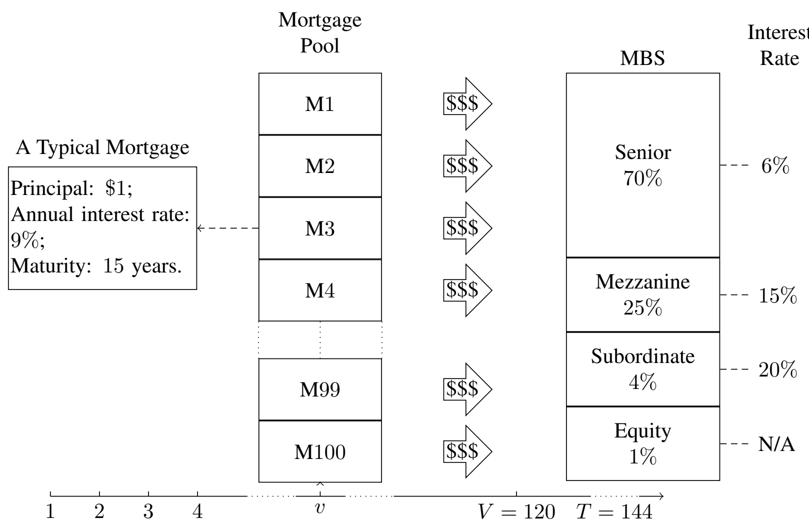
In this section we derive a relation between the loss distribution in a cashflow CDO and the interest rates paid by the tranches. We make the simplifying assumption that all losses and payments occur at the end of one period. This assumption is not unreasonable for subprime mortgages that have a short interest-rate reset period, which we take effectively as the lifetime of the mortgage (at the end of which it either pays back in full with interest or defaults). We refer to the constituents of the portfolio as "loans", though they could be other instruments. Figure 2 illustrates the structure of the portfolio and cashflows. As pointed out by [10, page xvii] there is "very little research or literature" available on cash CDOs; the complex waterfall structures that govern cashflows of such CDOs are difficult to model in a mathematically sound way. For technical descriptions of cashflow waterfall structures, see [23, Chapter 14].

Consider a portfolio of N loans, each with face value of one unit. Let $S(T)$ be the cash inflow from the investments made by the portfolio at time T, the end of the investment period. Next consider investors named $1, 2, \ldots, M$, with investor j investing amount I_j. The most senior investor, labeled 1, receives an interest rate r_1 (return per unit investment over the full investment period) if at all possible; this investor's cash inflow at time T is

$$Y_1(T) = \min\left\{S(T), (1 + r_1) I_1\right\}. \tag{4.1}$$

Proceeding in this way, investor j has payoff

FIGURE 2. **Illustration of schematic structure of MBS**



$$Y_j(T) = \min \left\{ S(T) - \sum_{1 \le i < j} Y_i(T),\ (1 + r_j) I_j \right\}. \tag{4.2}$$

Using the market pricing measure (risk-neutral measure) $\mathbb{Q}$ we should have

$$\mathbb{E}^{\mathbb{Q}}[Y_j(T)] = (1 + R_0) I_j, \tag{4.3}$$

where R_0 is the risk-free interest rate for the period of investment.

Given a model for $S(T)$, the rates r_j can be worked out, in principle, recursively from equation (4.2) as follows. Using the distribution of $X(1)$ we can back out the value of the supersenior rate r_1 from

$$\mathbb{E}^{\mathbb{Q}}\left[\min\left\{S(T), (1 + r_1) I_1\right\}\right] = \mathbb{E}^{\mathbb{Q}}[Y_1(T)] = (1 + R_0) I_1. \tag{4.4}$$

Now we use this value of r_1 in the equation for $Y_2(T)$:

$$\mathbb{E}^{\mathbb{Q}}\left[\min\left\{S(T) - Y_1(T), (1 + r_2) I_2\right\}\right] = \mathbb{E}^{\mathbb{Q}}[Y_2(T)] = (1 + R_0) I_2, \tag{4.5}$$

and (numerically) invert this to obtain the value of r_2 implied by the market model. Note that in equation (4.5) the random variable $Y_1(T)$ on the left is given by equation (4.1) using the already computed value of r_1. Proceeding in this way yields the full spectrum of tranche rates r_j.

Now we turn to a continuum model for tranches, again with one time period. Consider an idealized securitization structure ABS. Investors are subordinatized by a seniority parameter $y \in [0, 1]$. An investor in a thin "tranchelet" $[y, y + \delta y]$ invests the amount δy and is promised an interest rate of $r(y)$ (return on unit investment for the entire investment period) if there is no default. In this section we consider only one time period, at the end of which the investment vehicle closes.

Thus, if there is sufficient return on the investment made by the ABS, a tranche $[a, b] \subset [0, 1]$ will be returned the amount

$$\int_a^b \big(1 + r(y)\big)\, \mathrm{d}y.$$

In particular, assuming that the total initial investment in the portfolio is normalized to one, the maximum promised possible return to all the investors is $\int_0^1 \big(1 + r(y)\big)\, \mathrm{d}y$. The *portfolio loss* is

$$L = \int_0^1 \big(1 + r(y)\big)\, \mathrm{d}y - S(T), \tag{4.6}$$

where $S(T)$ is the total cash inflow, all assumed to occur at time T, from investments made by the ABS. Note that L is a random variable, since $S(T)$ is random.

Consider a thin tranche $[y, y + \delta y]$. If $S(T)$ is greater than the maximum amount promised to investors in the tranche $[y, 1]$, that is if

$$S(T) > \int_y^1 \big(1 + r(s)\big)\, \mathrm{d}s, \tag{4.7}$$

then the tranche $[y, y + \delta y]$ receives its maximum promised amount $\big(1 + r(y)\big)\, \delta y$. (If $S(T)$ is insufficient to cover the more senior investors, the tranchelet $[y, y + \delta y]$ receives nothing.) The condition (4.7) is equivalent to

$$L < \int_0^y \big(1 + r(s)\big)\, \mathrm{d}s, \tag{4.8}$$

as can be seen from the relation (4.6). Thus, the thin tranche receives the amount

$$\mathbb{1}_{\left[L < \int_0^y \left(1 + r(s)\right)\, \mathrm{d}s\right]} \big(1 + r(y)\big)\, \delta y$$

Using the risk-neutral probability measure $\mathbb{Q}$, we have then

$$\mathbb{Q}\left[L < \int_0^y \big(1 + r(s)\big)\, \mathrm{d}s\right] \big(1 + r(y)\big)\, \delta y = (1 + R_0)\, \delta y, \tag{4.9}$$

where R_0 is the risk-free interest rate for the period of investment. Thus,

$$\big(1 + r(y)\big) F_L\left(\int_0^y \big(1 + r(s)\big)\, \mathrm{d}s\right) = 1 + R_0 \tag{4.10}$$

where F_L is the distribution function of the loss L with respect to the measure $\mathbb{Q}$.

Let $\lambda(\cdot)$ be the function given by

$$\lambda(y) = \int_0^y \big(1 + r(s)\big)\, \mathrm{d}s, \tag{4.11}$$

which is strictly increasing as a function of y, with slope > 1 (assuming the rates $r(\cdot)$ are positive). Hence $\lambda(\cdot)$ is invertible. Then the loss distribution function is obtained as

$$F_L(l) = \frac{1 + R_0}{1 + r\big(\lambda^{-1}(l)\big)}. \tag{4.12}$$

If $r(y)$ are the market rates then the market-implied loss distribution function F_L is given by (4.12). On the other hand, if we have a prior model for the loss distribution F_L then the implied rates $r(y)$ can be computed numerically using (4.12).

A real tranche is a "thick" segment $[a, b] \subset [0, 1]$ and offers investors some rate $r_{[a,b]}$. This rate could be viewed as obtained from the balance equation:

$$(1 + r_{[a,b]})(b - a) = \int_a^b \big(1 + r(y)\big)\,\mathrm{d}y,$$

which means that the tranche rate is the average of the rates over the tranche:

$$r_{[a,b]} = \frac{1}{b-a}\int_a^b r(y)\,\mathrm{d}y. \tag{4.13}$$

5. Modeling Temporal Correlation in Subprime Securitization

We turn now to the study of a portfolio consisting of several CDOs (each homogeneous) belonging to different vintages. We model the loss by a "multi-stage" copula, one operating within each CDO and the other across the different CDOs. The motivation comes from the subprime context. Each CDO is comprised of subprime mortgages of a certain vintage, all with a common default/no-default decision horizon (typically two years). It is important to note that we *do not compare losses at different times* for the same CDO; we thus avoid problems in using a copula model across different time horizons.

Definition 5.1 (Vintage Correlation)**.** Suppose we have a pool of mortgages created at each time $v = 1, 2, \cdots, V$. Denote the default rates of each vintage observed at a fixed time $T > V$ as $p_1, p_2, \cdots, p_V$, respectively. We define *vintage correlation* $\phi_j := \mathrm{Corr}(p_1, p_j)$ for $j = 2, 3, \cdots, V$ 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 use the Gaussian copula approach of Li [34], widely used in industry to model default correlation across names. 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; monographs include [8, 18, 32, 40, 44]. Other works include, for example, [1, 2, 5, 6, 10, 17, 22, 24, 30, 33, 35, 45, 46]. 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. The paper [31] extends 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 [28] and in [47]. The work [11] provides a comparison of common CDO pricing models.

In our model 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

$$F_v(s) = \mathbb{P}[\tau_{v,i} < s],\ \forall i = 1, 2, ..., N, \tag{5.1}$$

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

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

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

$$C\left[F_1(x_1), F_2(x_2), ..., F_N(x_N)\right] = G(x_1, x_2, ..., x_N) \tag{5.2}$$

is a multivariate distribution function with marginals given by the distribution functions $F_1(x_1)$, $F_2(x_2)$,..., $F_N(x_N)$. Sklar [48] proved the converse, showing that for an arbitrary multivariate distribution function $G(x_1, x_2, ..., x_N)$ with continuous marginal distributions functions $F_1(x_1)$, $F_2(x_2)$,..., $F_N(x_N)$, there exists a unique C such that equation (5.2) holds. Therefore, in the case of default times, there is a C_v for each vintage v such that

$$C_v\left[F_v(t_1), F_v(t_2), ..., F_v(t_N)\right] = G_v(t_1, t_2, ..., t_N), \tag{5.3}$$

where G_v on the right is the joint distribution function of $(\tau_{v,1}, \ldots, \tau_{v,N})$. Since we assume F_v to be continuous and strictly increasing, we can find a standard Gaussian random variable $X_{v,i}$ such that

$$\Phi(X_{v,i}) = F_v(\tau_{v,i}) \quad \forall v = 1, 2, ..., V;\ i = 1, 2, ..., N, \tag{5.4}$$

or equivalently,

$$\tau_{v,i} = F_v^{-1}(\Phi(X_{v,i})) \quad \forall v = 1, 2, ..., V;\ i = 1, 2, ..., N, \tag{5.5}$$

where Φ is the standard normal distribution function. To see that this is correct, observe that

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

The Gaussian copula approach assumes that the joint distribution of $(X_{v,1}, \ldots, X_{v,N})$ is a multivariate normal distribution function Φ_N. 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 ρ_v for each vintage v. Consider the following definition

$$X_{v,i} := \sqrt{\rho_v} Z_v + \sqrt{1-\rho_v}\varepsilon_i \quad \forall i = 1, 2, \ldots, N;\ v = 1, 2, \ldots, V, \tag{5.6}$$

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 Φ_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 (5.6). Second, from equation (5.5) a set of N default times $\tau_{v,i}$ is obtained, which has the desired joint distribution function G_v. In equation (5.6), the common factor Z_v can be viewed as a latent variable that captures the default risk in the economy, and ε_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 ρ_v is the correlation between any two individual state variables. It is obvious that the higher the value of ρ_v, the greater the correlation between the default times of different mortgages.

Assume that we have a pool of N mortgages $i = 1, \ldots, N$ for each vintage $v = 1, \ldots, V$. 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 $[v, v']$. From equation (3.1), we can deduce that

$$Y_{v,i} := \log P_{v',i} - \log P_{v,i} = \Delta C_v + e_{v,i}, \tag{5.7}$$

where $\Delta C_v := \log C_{v'} - \log C_v$ is the change in the logarithm of a housing market index C_v, and $e_{v,i}$ are i.i.d. normal random variables for all $i = 1, 2, ..., N$, and $v = 1, 2, ..., V$. 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.

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

$$Z'_v := \frac{\Delta C_v}{\sigma_{\Delta C}}, \tag{5.8}$$

where $\sigma_{\Delta C}$ is the unconditional standard deviation of ΔC_v. Then we have

$$Y_{v,i} = Z'_v \sigma_{\Delta C} + e_{v,i}.$$

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

$$X'_{v,i} := \frac{Y_{v,i}}{\sigma_Y} = \frac{Z'_v \sigma_{\Delta C} + e_{v,i}}{\sigma_Y} = \frac{\sigma_{\Delta C}}{\sqrt{\sigma^2_{\Delta C} + \sigma^2_e}} Z'_v + \frac{\sigma_e}{\sqrt{\sigma^2_{\Delta C} + \sigma^2_e}} \varepsilon'_{v,i}$$

where σ_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

$$\sigma_Y = \sqrt{\sigma^2_{\Delta C} + \sigma^2_e}.$$

Define

$$\rho' := \frac{\sigma^2_{\Delta C}}{\sigma^2_{\Delta C} + \sigma^2_e}.$$

Then

$$X'_{v,i} = \sqrt{\rho'} Z'_v + \sqrt{1 - \rho'} \varepsilon'_{v,i} \quad \forall i = 1, 2, \ldots, N; \quad t = 1, 2, \ldots, T \tag{5.9}$$

Note that equation (5.9) has exactly the same form as equation (5.6). The default event is defined as $X'_{v,i} < X^{*\prime}$ where

$$X^{*\prime} := \frac{Y^*}{\sqrt{\sigma^2_{\Delta C} + \sigma^2_e}}.$$

Let

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

and

$$\tau^{*\prime}_v := F_v^{-1} \left(\Phi(X^{*\prime}) \right),$$

then the default event can be defined equivalently as $\tau'_{v,i} \leq \tau^{*\prime}$. The comparison between equation (5.9) and (5.6) 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. This is consistent with our remarks in the context of (3.2) that the Case-Shiller model provides a direct justification for using the Gaussian copula, with common risk factor being the housing 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,

$$\Delta C_v = \int_v^{v'} \mathrm{d} \log I_\tau,$$

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

$$\Delta C_v = \sum_{\tau \in [v, v']} (\log I_\tau - \log I_{\tau - 1}), \tag{5.10}$$

where the unit of τ is a quarter. If we model this index by some random shock arriving each quarter, equation (5.10) is a moving average process. Therefore, from equation (5.8) we know that Z'_v has positive serial correlation. Figure 1 shows that the time series of

Case-Shiller index changes exhibits strong autocorrelation, and is possibly integrated of order one.

6. The 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. We specify a time series model for the common risk factor in the Gaussian copula and determine the relationship between the serial correlation of the default rates and that of the common risk factor.

Proposition 6.1 (Default Probabilities and Numbers of Defaults)**.** *Let $k = 1, 2, ..., N$,*

$$X_k = \sqrt{\rho} Z + \sqrt{1-\rho}\,\varepsilon_k, \qquad \textit{and} \qquad X'_k = \sqrt{\rho'} Z' + \sqrt{1-\rho'}\,\varepsilon'_k \tag{6.1}$$

with

$$Z' = \phi Z + \sqrt{1-\phi^2}\,u, \tag{6.2}$$

where $\rho, \rho' \in (0,1)$, $\phi \in (-1,1)$, and Z, $\varepsilon_1, ..., \varepsilon_N$, $\varepsilon'_1, ..., \varepsilon'_N$, 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'_*:

$$A = \sum_{k=1}^{N} \mathbb{1}_{\{X_k \le X_*\}}, \qquad \textit{and} \qquad A' = \sum_{k=1}^{N} \mathbb{1}_{\{X'_k \le X'_*\}}, \tag{6.3}$$

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

$$\mathrm{Cov}(A, A') = N^2 \mathrm{Cov}(p, p'), \tag{6.4}$$

where

$$\begin{aligned} p &= p(Z) := \mathbb{P}[X_k \le X_* \,|\, Z] = \Phi\left(\frac{X_* - \sqrt{\rho} Z}{\sqrt{1-\rho}}\right), \qquad \textit{and} \\ p' &= \mathbb{P}[X'_k \le X'_* \,|\, Z'] = p'(Z'). \end{aligned} \tag{6.5}$$

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

Proof. We first show that

$$\mathbb{E}[AA'] = \mathbb{E}\left[\mathbb{E}[A \,|\, Z]\mathbb{E}[A' \,|\, Z']\right]. \tag{6.6}$$

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

$$\begin{aligned} &\mathbb{E}[f(Z)g(Z')F(Z,\boldsymbol{\epsilon})G(Z',\boldsymbol{\epsilon'})] \\ &= \int f(z)g(\underbrace{\phi z + \sqrt{1-\phi^2}x}_{z'})F(z,\underbrace{y_1, ..., y_N}_{\boldsymbol{y}})G(z',\underbrace{y'_1, ..., y'_N}_{\boldsymbol{y'}})\ \mathrm{d}\Phi(z,x,\boldsymbol{y},\boldsymbol{y'}) \\ &= \int f(z)g(z')\left[\left\{\int F(z,\boldsymbol{y})\ \mathrm{d}\Phi(\boldsymbol{y})\right\}\left\{\int G(z',\boldsymbol{y'})\ \mathrm{d}\Phi(\boldsymbol{y'})\right\}\right]\ \mathrm{d}\Phi(z,x) \\ &= \mathbb{E}\left[f(Z)g(Z')\mathbb{E}[F(Z,\boldsymbol{\epsilon}) \,|\, Z]\mathbb{E}[G(Z',\boldsymbol{\epsilon'}) \,|\, Z']\right]. \end{aligned} \tag{6.7}$$

This says that

$$\mathbb{E}\left[F(Z,\varepsilon)G(Z',\varepsilon') \mid Z, Z'\right] = \mathbb{E}[F(Z,\varepsilon) \mid Z]\mathbb{E}[G(Z',\varepsilon') \mid Z']. \tag{6.8}$$

Taking expectation on both sides of equation (6.8) with respect to Z and Z', we obtain

$$\mathbb{E}\left[F(Z,\varepsilon)G(Z',\varepsilon')\right] = \mathbb{E}\left[\mathbb{E}[F(Z,\varepsilon) \mid Z]\mathbb{E}[G(Z',\varepsilon') \mid Z']\right]. \tag{6.9}$$

Substituting $F(Z,\varepsilon) = A$, and $G(Z',\varepsilon') = A'$, we have equation (6.6) and

$$\begin{aligned} \mathbb{E}[AA'] &= \mathbb{E}\left[\mathbb{E}[A \mid Z]\mathbb{E}[A' \mid Z']\right] \\ &= \mathbb{E}[NpNp'] = N^2\mathbb{E}[pp'], \end{aligned} \tag{6.10}$$

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 (6.9) again with $F(Z,\varepsilon) = A$, and $G(Z',\varepsilon') = 1$, or indeed, much more directly by repeated expectations, we have

$$\mathbb{E}[A] = N\mathbb{E}[p], \qquad \text{and} \qquad \mathbb{E}[A'] = N\mathbb{E}[p']. \tag{6.11}$$

Hence we conclude that

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

We have

$$\mathrm{Var}(A) = \mathbb{E}\left[\mathbb{E}[A^2 \mid Z]\right] - N^2(\mathbb{E}[p])^2 = N\mathbb{E}[p(1-p)] + N^2\,\mathrm{Var}(p). \tag{6.12}$$

Similarly,

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

Putting everything together, we have for the correlations:

$$\begin{aligned} \mathrm{Corr}(A, A') &= \frac{\mathrm{Corr}(p, p')}{\sqrt{1 + \frac{\mathbb{E}[p(1-p)]}{N\,\mathrm{Var}(p')}}\sqrt{1 + \frac{\mathbb{E}[p'(1-p')]}{N\,\mathrm{Var}(p)}}} \\ &= \mathrm{Corr}(p, p') \quad \text{as} \quad N \to \infty. \end{aligned} \tag{6.13}$$

□

Theorem 6.2 (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* (5.1)*,* (5.5)*, and* (5.6) *with common risk factor Z_v being a zero-mean stationary Gaussian process. Assume further that $\rho_v = \mathrm{Corr}(X_{v,i}, X_{v,j})$, 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 $\phi_{v,v'} = \mathrm{Corr}(Z_v, Z_{v'}) \neq 0$, where Z_v is the common Gaussian risk factor process. Moreover, in the large portfolio limit, $\mathrm{Corr}(A_v, A_{v'})$ approaches a limiting value determined by $\phi_{v,v'}$, ρ_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,

$$\mathbb{P}(A_v = k|Z_v) = \binom{N}{k} p_v^k (1-p_v)^{N-k} \tag{6.14}$$

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

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

with

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

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

$$p_v = \mathbb{P}\left(X_{v,i} \le X_v^*|Z_v\right) = \Phi\left(Z_v^*\right), \tag{6.15}$$

where

$$Z_v^* = \frac{X_v^* - \sqrt{\rho_v} Z_v}{\sqrt{1-\rho_v}}. \tag{6.16}$$

Similarly,

$$p_{v'} = \Phi\left(Z_{v'}^*\right), \tag{6.17}$$

where

$$Z_{v'}^* = \frac{X_{v'}^* - \sqrt{\rho_{v'}} Z_{v'}}{\sqrt{1-\rho_{v'}}}. \tag{6.18}$$

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

$$Z_v = \phi_{v,v'} Z_{v'} + \sqrt{1-\phi_{v,v'}^2}\, u_{v,v'} \qquad \text{for } t > j, \tag{6.19}$$

where $u_{v,v'}$ are standard Gaussians that are independent of $Z_{v'}$. Combining equation (6.16), (6.18) and (6.19), we have

$$Z_v^* = a\phi_{v,v'} Z_{v'}^* + \frac{X_t^* - b\phi_j X_{v'}^*}{\sqrt{1-\rho_t}} - \frac{\sqrt{\rho_v(1-\phi_{v,v'}^2)}}{\sqrt{1-\rho_v}} u_{v,v'}, \tag{6.20}$$

where

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

$$\begin{aligned} \mathrm{Cov}(p_v, p_{v'}) &= \mathrm{Cov}\left(\Phi(Z_v^*), \Phi(Z_{v'}^*)\right) \\ &= \mathrm{Cov}\left(\Phi\left(a\phi_{v,v'} Z_{v'}^* + \frac{X_v^* - b\phi_{v,v'} X_{v'}^*}{\sqrt{1-\rho_v}} - \frac{\sqrt{\rho_v(1-\phi_{v,v'}^2)}}{\sqrt{1-\rho_v}} u_{v,v'}\right), \Phi\left(Z_{v'}^*\right)\right). \end{aligned} \tag{6.21}$$

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 $\phi_{v,v'}$, ρ_v, and $\rho_{v'}$. They are nonzero if and only if $\phi_{v,v'} \neq 0$. Applying Proposition 6.1, we know that

$$\mathrm{Corr}(A_v, A_{v'}) = \frac{\mathrm{Corr}(p_v, p_{v'})}{\sqrt{1 + \frac{\mathbb{E}[p_v(1-p_v)]}{N\,\mathrm{Var}(p_{v'})}}\sqrt{1 + \frac{\mathbb{E}[p_{v'}(1-p_{v'})]}{N\,\mathrm{Var}(p_v)}}} \quad \forall v \neq v'. \tag{6.22}$$

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

Equations (6.21) and (6.22) 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 (6.21) how the vintage correlation of default rates depends on $\phi_{v,v'}$. The theorem below (whose proof extends an idea from [37]) shows that this dependence is always positive.

Theorem 6.3 (Dependence on Common Risk Factor)**.** *Under the same settings as in Theorem 6.2, assume that both the serial correlation $\phi_{v,v'}$ of the common risk factor and the individual state variable correlation ρ_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 6.1. 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

$$\begin{aligned}\operatorname{Cov}(A_v, A_{v'}) &= \sum_{k,l=1}^{N} \operatorname{Cov}(\mathbb{1}_{[X_{v,k} \leq X_v^*]}, \mathbb{1}_{[X_{v',l} \leq X_{v'}^*]}) \\ &= N^2 \operatorname{Cov}(\mathbb{1}_{[X \leq X_v^*]}, \mathbb{1}_{[X' \leq X_{v'}^*]}),\end{aligned} \tag{6.23}$$

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

$$\operatorname{Cov}(X, X') = \phi_j \sqrt{\rho_v \rho_{v'}}. \tag{6.24}$$

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

$$\begin{aligned} W_1 &:= X, \\ W_2 &:= \frac{1}{\sqrt{1 - \rho_v \rho_{v'} \phi_{v,v'}^2}} [X' - \phi_{v,v'} \sqrt{\rho_v \rho_{v'}} X]. \end{aligned} \tag{6.25}$$

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

$$\begin{aligned} X &= W_1, \\ X' &= \phi_{v,v'} \sqrt{\rho_v \rho_{v'}} W_1 + \sqrt{1 - \rho_v \rho_{v'} \phi_{v,v'}^2}\, W_2. \end{aligned} \tag{6.26}$$

Let

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

The assumption that ρ 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

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

where $\varphi(\cdot)$ is the probability density function of the standard normal distribution. The third equality follows from equation (6.9). The fifth equality follows from equation (6.26). The unconditional expectation of p_v is independent of α, because

$$\mathbb{E}(p_v) = \mathbb{E}\left[\,\mathbb{P}(X_{v,i}\le X_v^* | Z_v)\,\right] = \Phi(X_v^*). \tag{6.27}$$

It follows that

$$\begin{aligned}
\frac{\partial}{\partial\alpha}\operatorname{Cov}(p_v, p_{v'}) &= \int_{-\infty}^{X_v^*} \varphi\left(\frac{X_{v'}^* - \alpha w_1}{\sqrt{1-\alpha^2}}\right)\varphi(w_1)\frac{\partial\left(\frac{X_{v'}^*-\alpha w_1}{\sqrt{1-\alpha^2}}\right)}{\partial\alpha}\,\mathrm{d}w_1 \\
&= \int_{-\infty}^{X_v^*} \varphi\left(\frac{X_{v'}^* - \alpha w_1}{\sqrt{1-\alpha^2}}\right)\varphi(w_1)\frac{-w_1+\alpha X_{v'}^*}{(1-\alpha^2)^{\frac{3}{2}}}\,\mathrm{d}w_1 \\
&= -\frac{1}{(1-\alpha^2)^{\frac{3}{2}}}\int_{-\infty}^{X_v^*}(w_1-\alpha X_{v'}^*)\varphi\left(\frac{X_{v'}^* - \alpha w_1}{\sqrt{1-\alpha^2}}\right)\varphi(w_1)\,\mathrm{d}w_1.
\end{aligned} \tag{6.28}$$

The last two terms in the integrand simplify to

$$\varphi\left(\frac{X_{v'}^* - \alpha w_1}{\sqrt{1-\alpha^2}}\right)\varphi(w_1) = \frac{1}{2\pi}\exp\left[-\frac{(w_1-\alpha X_{v'}^*)^2 + {X_{v'}^*}^2(1-\alpha^2)}{2(1-\alpha^2)}\right]. \tag{6.29}$$

Substituting equation (6.29) into (6.28), we have

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

Make a change of variable and let

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

It follows, upon further simplification, that

$$\frac{\partial}{\partial\alpha}\operatorname{Cov}(p_v, p_{v'}) = \frac{1}{2\pi\sqrt{1-\alpha^2}}\exp\left(-\frac{X_v^{*2} - 2\alpha X_v^* X_{v'}^* + {X_{v'}^*}^2}{2(1-\alpha^2)}\right) > 0. \tag{6.30}$$

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

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

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

with ρ_s and $\phi_{v,v'}$ assumed to be positive, we know that the partial derivatives of the covariance with respect to $\phi_{v,v'}$, ρ_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 ρ_s), which can be seen from equation (6.15). It follows that the serial correlation of p_v has positive partial derivative with respect to $\phi_{v,v'}$. Recall equation (6.21), which shows that the covariance of p_v and $p_{v'}$ is zero for any value of ρ_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 (6.22), 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 $\phi_{v,v'}$. $\square$

7. Monte Carlo Simulations

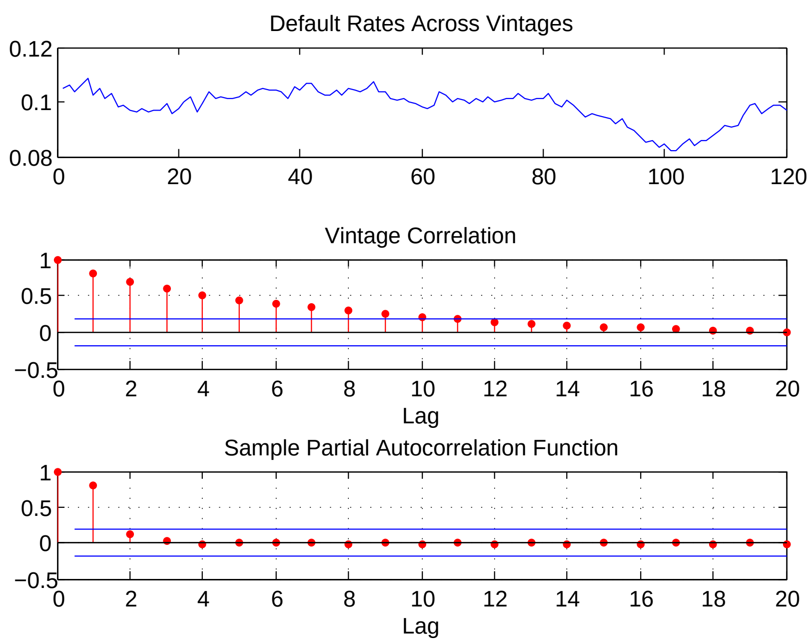
In this section, we study the link between serial correlation in a common risk factor and vintage correlation in pools of mortgages in two sets of simulations: First, a series of mortgage pools is simulated to illustrate the analytical results of Section 6. Second, a waterfall structure is simulated to study temporal correlation in MBS.

7.1. 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 (5.1), (5.5), and (5.6). 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 $v = 1, 2, \ldots, 120$. 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,

$$Z_v = \phi Z_{v-1} + \sqrt{1-\phi^2}\, u_v \quad \forall v = 2, 3, \ldots, 120. \tag{7.1}$$

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 $\phi = 0.95$. 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 (5.6). The error ε_i in equation (5.6) is independent of u_v.

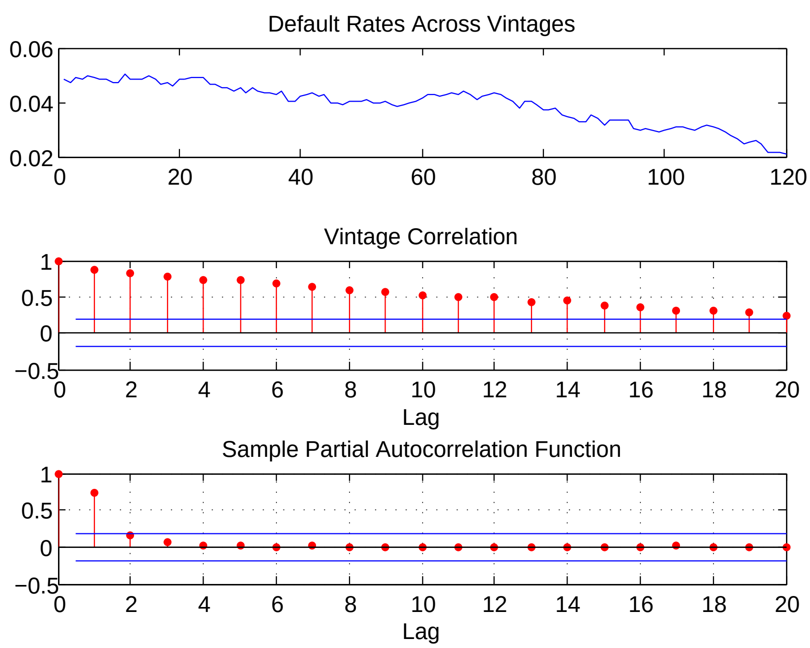
FIGURE 3. **Serial Correlation in Default Rates of Subprime Mortgages**



To simulate the actual default rates of mortgages, we need to specify the marginal distribution functions of default times $F(\cdot)$ as in equation (5.1). 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 $S = 1, 2, \ldots$. It is easy to verify that $F_v(T) = \mathcal{F}(T - v)$. The values of the function $\mathcal{F}(\cdot)$ are specified 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 1998 to 2007, 100 mortgages are created.

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.

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



We construct a time series $\tau_{v,i}$ of default times of mortgage i issued at time v according to equation (5.5). (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.) 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, $\tau_v^* = 24$; 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 3. For the prime case, average simulated default rates are plotted in Figure 4. Note that because of the decreasing time window to default, the default rates in Figure 4 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 3 and 4 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 autocorrelation functions are plotted in the third rows of Figures 3 and 4. They are significant only at lag one. This phenomenon is also observed when we set ϕ 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 (6.15) at $Z_v^* = 0$ to first order, we have

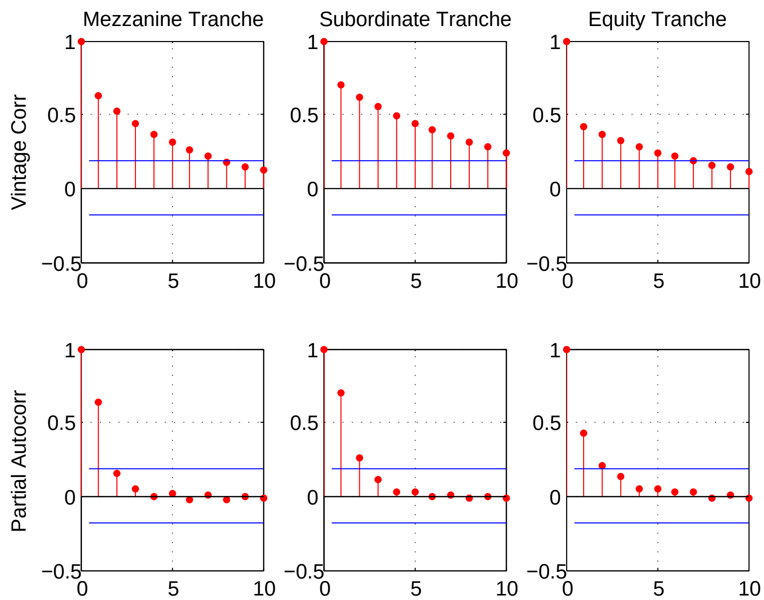
$$p_v \approx \frac{1}{2} + \frac{1}{\sqrt{2\pi}} Z_v^*. \tag{7.2}$$

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

7.2. Vintage Correlation in Waterfall Structures. We have already shown using the Gaussian copula approach 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: (i) A time line of 120 months and an observation time $T = 144$. (ii) The mortgage contract has a principal of \$ 1, maturity of 15 years, and annual interest rate 9%. Fixed monthly payments are received until the mortgage defaults or is paid in full. A pool of 100 such mortgages is created every month. (iii) There is a pool of 100 units of MBS, each of principal \$1, securitized from each month's mortgage cohort. There are four tranches: 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; 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 2. 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 (5.1), (5.5), and (5.6). 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 ρ is set to be 0.5. The unconditional default probabilities over time are obtained from Table 1.

FIGURE 5. **Serial Correlation in Principal Losses of Subprime MBS**



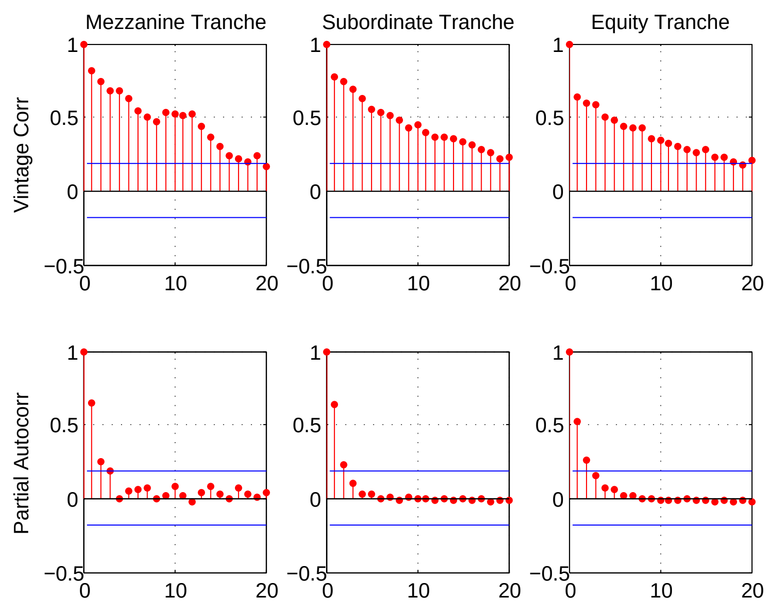
The first row plots the vintage correlation of the principal loss of each tranche. The correlation is estimated using the sample autocorrelation function. The second row plots the partial autocorrelation functions.

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.

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 5 and 6 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.

The series of present values of cash flows for each tranche and their sample autocorrelation and partial autocorrelation functions are plotted in Figures 7 and 8 for subprime

FIGURE 6. **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.

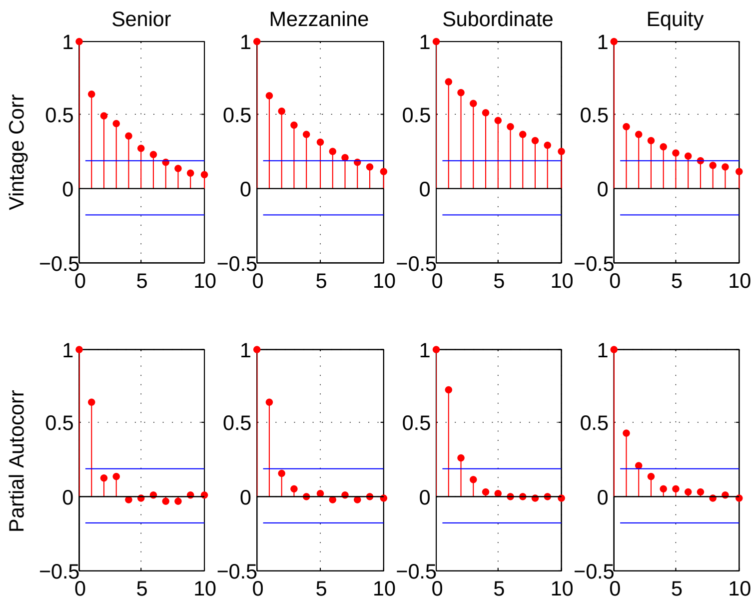
and prime scenarios, respectively. The senior tranche displays a significant first-order autocorrelation coefficient due to losses in interest payments although there are no losses in principal. The partial autocorrelation functions, which have significant positive values for more than one lag, suggest 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.

Acknowledgment. Discussions with Darrell Duffie and Jean-Pierre Fouque helped improve the paper. We thank seminar and conference participants at Stanford University, LSU, UC Merced, UC Riverside, UC Santa Barbara, the 2010 Midwest Econometrics Group Meetings in St. Louis, Maastricht University, and CREATES. EH acknowledges support from the Danish National Research Foundation. ANS acknowledges research support received as a Mercator Guest Professor at the University of Bonn in 2011 and 2012, and discussions with Sergio Albeverio and Claas Becker.

References

1. Andersen, L. and J. Sidenius: Extensions to the Gaussian Copula: Random Recovery and Random Factor Loadings, *Journal of Credit Risk* **1**, 1 (2005) 29–70.
2. Andersen, L., Sidenius, J. and Basu, S.: All your Hedges in One Basket, *Risk* (2003) 67–72.

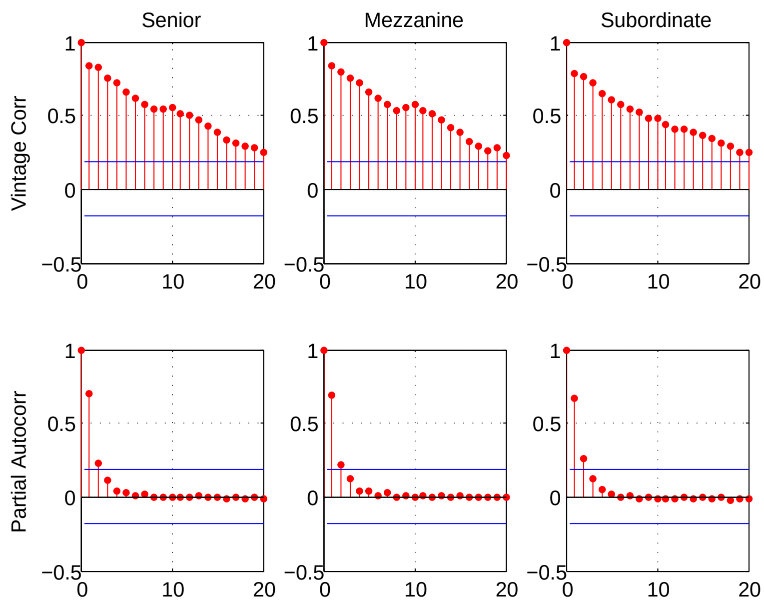
FIGURE 7. **Cashflow correlation in Subprime MBS**



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.

3. Bailey, M. J., Muth, R. F. and Nourse, H. O.: A regression method for real estate price index construction, *Journal of the American Statistical Association* **58**, 304 (1963) 933–942.
4. Bajari, P., Chu, C. and Park, M.: An Empirical Model of Subprime Mortgage Default From 2000 to 2007, Working Paper 14358, NBER, 2008.
5. Beare, B.: Copulas and Temporal Dependence, *Econometrica* **78**, 1 (2010) 395–410.
6. Berd, A., Engle, R. and Voronov, A.: The Underlying Dynamics of Credit Correlations, *Journal of Credit Risk* **3**, 2 (2007) 27–62.
7. Bhardwaj, G. and Sengupta, R.: Subprime Mortgage Design, Working Paper 2008-039E, Federal Reserve Bank of St. Louis, 2011.
8. Bluhm, C., Overbeck, L. and Wagner, C.: *An Introduction to Credit Risk Modeling*, Chapman & Hall/CRC, London, 2002.
9. Bogue, A. G.: Land Credit for Northern Farmers 1789-1940, *Agricultural History* **50**, 1 (1976) 68–100.
10. Brigo, D., Pallavacini, A. and Torresetti, R.: *Credit Models and the Crisis: A Journey into CDOs, Copulas, Correlations and Dynamic Models*, Wiley, Hoboken, 2010.
11. Burtschell, X., Gregory, J. and Laurent, J.-P.: A Comparative Analysis of CDO Pricing Models, *Journal of Derivatives* **16**, 4 (2009) 9–37.
12. Caballero, R. J. and Krishnamurthy, A.: Global Imbalances and Financial Fragility, *American Economic Review* **99**, 2 (2009) 584–588.
13. Case, K. E. and Shiller, R. J.: Prices of Single Family Homes Since 1970: New Indexes for Four Cities, *New England Economic Review* (1987) 45–56.
14. Chomsisengphet, S. and Pennington-Cross, A. The Evolution of the Subprime Mortgage Market, Discussion Paper 1, Federal Reserve Bank of St. Louis Review, January/February 2006.
15. Crouhy, M. G., Jarrow, R. A. and Turnbull, S. M.: The Subprime Credit Crisis of 2007, *Journal of Derivatives* **16**, 1 (2008) 81–110.

FIGURE 8. **Cashflow correlation in 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.

16. Daglish, T.: What Motivates a Subprime Borrower to Default?, *Journal of Banking and Finance* **33**, 4 (2009) 681–693.
17. Das, S. R., Freed, L., Geng, G. and Kapadia, N.: Correlated Default Risk, *Journal of Fixed Income* **16**, 2 (2006) 7–32.
18. Duffie, D. and Singleton, K.: *Credit Risk: Pricing, Measurement, and Management*, Princeton University Press, Princeton, 2003.
19. Fabozzi, F. J.: *The Handbook of Mortgage-Backed Securities*, Probus Publishing Company, Chicago, 1988.
20. Figlewski, S.: Viewing the Crisis from 20,000 Feet Up, *Journal of Derivatives* **16**, 3 (2009) 53–61.
21. Frees, E. W. and Valdez, E. A.: Understanding Relationships Using Copulas, *North American Actuarial Journal* **2**, 1 (1998) 1–25.
22. Frey, R. and McNeil, A.: Dependent Defaults in Models of Portfolio Credit Risk, *Journal of Risk* **6**, 1 (2003) 59–92.
23. Garcia, J. and Goosens, S.: *The Art of Credit Derivatives: Demystifying the Black Swan*, Wiley, Hoboken, 2010.
24. Giesecke, K.: A Simple Exponential Model for Dependent Defaults, *Journal of Fixed Income* **13**, 3 (2003) 74–83.
25. Goetzmann, W. N.: The Accuracy of Real Estate Indices: Repeat Sale Estimators, *Journal of Real Estate Finance and Economics* **5**, 3 (1992) 5–53.
26. Gorton, G.:The Panic of 2007, Working Paper 14358, NBER, 2008.
27. Gorton, G.: Information, Liquidity, and the (Ongoing) Panic of 2007, *American Economic Review* **99**, 2 (2009) 567–572.
28. Graziano, G. D. and Rogers, L. C. G.: A Dynamic Approach to the Modeling of Correlation Credit Derivatives Using Markov Chains, *International Journal of Theoretical and Applied Finance* **12**, 1 (2009) 45–62.

29. Hayre, L. S., Saraf, M., Young, R. and Chen, J.: Modeling of Mortgage Defaults, *Journal of Fixed Income* **17**, 4 (2008) 6–30.
30. Hull, J., Predescu, M. and White, A.: The Valuation of Correlation-Dependent Credit Derivatives Using a Structural Model, *Journal of Credit Risk* **6**, 3 (2010) 53–60.
31. Hull, J. and White, A.: Valuing Credit Default Swaps II: Modeling Default Correlations, *Journal of Derivatives* **8**, 3 (2001) 12–22.
32. Lando, D.: *Credit Risk Modeling: Theory and Applications*, Princeton University Press, Princeton, 2004.
33. Laurent, J. and Gregory, J.: Basket Default Swaps, CDOs and Factor Copulas, *Journal of Risk* **7**, 4 (2005) 103–122.
34. Li, D.: On Default Correlation: A Copula Function Approach, *Journal of Fixed Income* **9**, 4 (2000) 43–54.
35. Lindskog, F. and McNeil, A.: Common Poisson Shock Models: Applications to Insurance and Credit Risk Modelling, *ASTIN Bulletin* **33**, 2 (2003) 209–238.
36. Lucas, D. J., Goodman, L. S. and Fabozzi, F. J.: *Collateralized Debt Obligations*, 2nd edition, Wiley, Hoboken, 2006.
37. Meng, C. and Sengupta, A. N.: CDO tranche sensitivities in the Gaussian Copula Model, *Communications on Stochastic Analysis* **5**, 2 (2011) 387–403.
38. Merton, R. C.: On the Pricing of Corporate Debt: The Risk Structure of Interest Rates, *Journal of Finance* **29**, 2 (1974) 449–470.
39. Murphy, D.: A Preliminary Enquiry into the Causes of the Credit Crunch, *Quantitative Finance* **8**, 5 (2008) 435–451.
40. O'Kane, D.: *Modelling Single-Name and Multi-Name Credit Derivatives*, Wiley, Hoboken, 2008.
41. Pavel, C.: *Securitization: The Analysis and Development of the Loan-Based/Asset-Based Securities Market*, Probus Pblishing, Chicago, 1989.
42. Reinhart, C. M. and Rogoff, K. S.: Is the US Sub-Prime Financial Crisis So Different? An International Historical Comparison, *American Economic Review* **98**, 2 (2008) 339–344.
43. Reinhart, C. M. and Rogoff, K. S.: The Aftermath of Financial Crises, *American Economic Review* **99**, 2 (2009) 466–472.
44. Schönbucher, P.: *Credit Derivatives Pricing Models: Models, Pricing and Implementation*, Wiley, Hoboken, 2003.
45. Schönbucher, P. and Schubert, D.: Copula Dependent Default Risk in Intensity Models, Working paper, Bonn University, 2001.
46. Servigny, A. and Renault, O.: Default Correlation: Empirical Evidence, Working paper, Standard and Poor's, 2002.
47. Sidenius, J., V. Piterbarg and Andersen, L.: A New Framework for Dynamic Credit Portfolio Loss Modelling, *International Journal of Theoretical and Applied Finance* **11**, 2 (2008) 163–197.
48. Sklar, A.: Fonctions de Repartition a n Dimensions et Leurs Marges, *Publ. Inst. Stat. Univ. Paris* **8** (1959) 229–231.
49. Wyngarden, H.: An Index of Local Real Estate Prices, *Michigan Business Studies* **1**, 2 (1927).

ERIC HILLEBRAND: CREATES - CENTER FOR RESEARCH IN ECONOMETRIC ANALYSIS OF TIME SERIES, DEPARTMENT OF ECONOMICS AND BUSINESS, AARHUS UNIVERSITY, DENMARK
E-mail address: `ehillebrand@creates.au.dk`

AMBAR N. SENGUPTA: DEPARTMENT OF MATHEMATICS, LOUISIANA STATE UNIVERSITY, BATON ROUGE, LA 70803, USA
E-mail address: `ambarnsg@gmail.com`

JUNYUE XU: FINANCIAL ENGINEERING PROGRAM, HAAS SCHOOL OF BUSINESS, UNIVERSITY OF CALIFORNIA, BERKELEY, CA 94720, USA
E-mail address: `junyue@gmail.com`