Intensity of default in sovereign bonds: Estimation of an unobservable process

Otero, Karina V.

The Pennsylvania State University

2016

Intensity of default in sovereign bonds:

Estimation of an unobservable process

Karina Otero[*]

February 2016

Abstract

This paper proposes a new approach to estimate general stationary diffusion processes that describe the evolution of unobserved arrival rates of credit events on sovereign bonds, allowing for arbitrary parametric drift and diffusion specifications. The solutions and transition processes for stationary diffusions are generally unknown in closed form and therefore standard maximum likelihood methods do not apply. Moreover, the arrival rates of credit events on sovereign bonds are unobservable and a direct nonparametric estimation does not work. This paper overcomes these challenges combining a semi-nonparametric estimator in the framework of the Efficient Method of Moments, Gallant and Tauchen (1996), and a reduced-form model for pricing sovereign bonds and credit default swaps. The application for Brazil sovereign assets explores the performance of the model under different specifications of the intensity process.

Keywords: Efficient Method of Moments (EMM), semi-nonparametric (SNP) econometrics, Hermite, latent variables, estimation of stochastic differential equations, estimation of diffusions, asset pricing, numerical methods for partial differential equations, credit risk, cox process, credit derivatives, credit default swaps (CDS).

JEL Classification: C14, C32, C58, C63, G12, G13.

I Introduction

Sovereign defaults have been modeled through a number of relevant factors among which are a complex trade-off of incentives, the impossibility of repayment, the composition of the debt and political decisions

1

to force a renegotiation of the debt, see Duffie, Pedersen, and Singleton (2003) for a review. Nevertheless, the implementation of a statistical methodology to determine the probability of a credit event in sovereign bonds faces the immediate obstacle of a natural lack of data on the triggers of the defaults and the default events themselves.[1]

Reduced-form models of sovereign default propose an alternative approach that is more suitable to the available information. In this framework, the default is modeled as an exogenous stopping time that might or not depend on additional exogenous observable processes. The timing of this stopping time follows an exogenous Poisson process, whose parameter is known as "the intensity of default" Duffie and Singleton (1999), Jarrow and Turnbull (1995), Lando (1998). In principle, if the intensity of default is given, then the probability of a credit event can be computed and the cash flows promised by a risky financial asset can be discounted accordingly.

Under the so called "doubly stochastic framework", the intensity is stochastic and follows a diffusion process. For instance, given some initial condition, the intensity λ_t can be modeled as the solution of a stochastic differential equation such as

$$d \ln \lambda_t = \kappa \left(\beta - \ln \lambda_t \right) dt + \sigma dW_t, \tag{1}$$

where W_t is a standard Brownian motion and (κ, β, σ) are parameters. The conditional density function of λ_t is known in closed form only for some particular specifications of the drift and the diffusion functions, for instance, the specification just given in equation (1) implies that λ_t has a log-normal distribution. In these cases, the likelihood function of the reduced-form model can be constructed and the parameters estimated by maximum likelihood; see for instance Pan and Singleton (2008).[2] Pan and Singleton (2008) estimate the parameters of equation (1) implicit in the term structures of sovereign CDS spreads by maximum likelihood and in order to avoid a deterioration in the fit of the model during periods of market turmoil, the implied likelihood function is modified by introducing pricing errors that follow a normal distribution with time-varying variances that depend on the bid/ask spread. Thus, the empirical evidence would suggest that equation (1) might not be the best model to fit the data.

This paper develops a reduced form model for the CDS prices similar to Pan and Singleton (2008) but assuming that the intensity of default follows a stationary diffusion process with more general specification. Thus, the density function of λ_t is unknown in closed form. Following the methodology of the Efficient Method of Moments (EMM) in Gallant and Tauchen (1996), the estimator of the parameters of the model is constructed with a semi-nonparametric (SNP) density function of an auxiliary model fitted to the data of interest, in this case the CDS and bond prices. Given the parameters values, the reduced-form model in this paper can price CDS and bonds using in principle any stationary diffusion processes for the implicit intensity of default. A by-product of this approach is that the modeling strategy for the CDS of different maturities can be the same, while Pan and Singleton (2008) assume that the 5-year CDS contract is perfectly priced and other maturities price with some error.[3] Moreover, the estimation strategy

---

[1] In general, the models developed to estimate the probabilities of credit events for corporate debt are not applicable to sovereign credit risk because the methods take advantage of large cross section databases with information about features of the debts and debtors.

[2] If the model is carefully constructed using particular parametric specifications for the diffusion process, the standard econometric techniques apply.

[3] In Pan and Singleton (2008) the 5-year CDS contract is assumed to be perfectly priced and it is inverted to compute the

in this paper does not rely on ad-hoc assumptions about the distribution of the pricing errors to estimate the parameters of interest.[4]

Among the literature valuing CDS contracts, Berndt, Douglas, Duffie, Ferguson, and Schranz (2008) for corporate contracts also assume that $\ln(\lambda_t)$ follows an Ornstein-Uhlenbeck process. Zhang (2004) for sovereign and Longstaff, Mithal, and Neis (2005) for corporate underlyings have assumed that λ_t follows a square-root diffusion which implies a scaled noncentral chi-square distribution for λ_t. The literature of term structure models shares some similar issues because it also deals with models based on diffusion processes, as an example, Ahn and Gao (1999) assumes that the state variable follows a "three-halves" diffusion, with an implied probability density that involves the modified Bessel function of first kind.

Duffie, Pedersen, and Singleton (2003) proposes a parametric model for the sovereign yield spreads of Russia in which the intensity of default is a component of the state variables and the state vector follows a multivariate affine diffusion process with specifications of type Gaussian and square-root. To estimate this model, Duffie, Pedersen, and Singleton (2003) introduces a new approximate maximum likelihood estimator that applies to a set of affine diffusions. The approximate likelihood is derived by partitioning the state vector and involves the mixture of a conditional noncentral chi-square, and a conditional normal.[5]

Besides the EMM, there are alternative methods for estimating parameters in continuous-time models. Quasi-maximum likelihood techniques can be used for some applications for instance, see Hurn, Lindsay, and McClelland (2013). The quasi-maximum estimators of the parameters of the structural model can be consistent (but inefficient), and when there are latent states they suffer from a higher risk of misspecification bias. Pan (2002) estimates a stochastic volatility model for option pricing by GMM. Her estimation strategy takes full advantage of the analytical tractability of the model and a set of moment conditions is derived from the joint conditional moment generating function of stock returns and volatility. This estimation strategy is not available if the model does not imply a moment generating function in closed form.

Pedersen (1995) derives a sequence of approximations to the unknown transition densities that incorporates the parametric functional forms of the drift and the diffusion coefficients. The sequence converges to the true densities and it is subsequently used to obtain a sequence of approximations to the true likelihood. These approximations are in general not available in closed form which increases the computational burden of estimating complex models. Johannes, Polson, and Stroud (2009) combine time-discretization schemes with Monte Carlo methods to estimate models with latent states. Their methodology approximates the true likelihood through time-discretization of the diffusion processes and some particular features of the models, such as the specifications of the distribution for the pricing errors.

Aït-Sahalia (2008) develops closed-form likelihood expansions for multivariate diffusions which can increase the accuracy and speed achieved by alternative methods used to approximate the log-likelihood of discretely observed diffusions, see Hurn, Jeisman, and Lindsay (2007), Jensen and Poulsen (2002) and Stramer and Yan (2007). These expansions are formulated as functions of the particular parametric specifications of the diffusion processes under consideration and they might not be applicable with unobserv-

---

unobservable initial conditions λ_t for all t. Here the assumption is not needed because there is an additional source of information that comes from the underlying bond prices.

[4]The only assumption is that the pricing errors have 0 mean.

[5]The normality property is obtained after conditioning on the entire path of a subset of the state variables

able latent states. Aït-Sahalia and Kimmel (2007) develop a method for maximum likelihood estimation of diffusion-based stochastic volatility models. Under unobservable state variables, the method requires a sufficiently tractable specification of the model to map observables into unobservable state variables.[6] Pastorello, Patilea, and Renault (2003) observe that in applications of Aït-sahalia (2002) to models with affine-type diffusion processes and latent variables, the direct optimization of the likelihood constructed using the expansions pushes the implied values of the unobservable factors towards their boundaries where the approximated likelihood is infinite. Pastorello, Patilea, and Renault (2003) develop a model where implied-state backfitting strategies of the likelihood maximization can correct this issue. This alternative implements an iteratively implied state methodology that requires the repeated computation of the inverse transformation between observables and unobservables for all the considered periods, it can be affected by the assumptions about the distribution of the pricing errors in small samples, see Garcia, Ghysels, and Renault 2010, and the method is not efficient. The model proposed here has unobservable latent states and it is highly nonlinear, thus in this case the closed-form expansions for multivariate diffusions does not seem to be the best choice to construct an estimator.

In view of the estimation strategies available so far, the complexity of estimating more general specifications for the intensity of default is in part due to the lack of a sufficiently simple closed form expression to model its transition density, and the fact that the intensity of default is an unobservable process. In the literature, this process is estimated using a structural or reduced form model which allows its identification. The relevant likelihood of the whole model is usually a transformation of the proposed approximations to the unknown transition densities, and both these transformations and the approximations are model specific and must be derived for each specification that the researcher intends to estimate. The complexity of deriving and computing this (in principle) intuitive construction could be one of the reasons why the literature has proposed estimators for the intensities of default that work only under the assumption that the intensity follows relatively simple diffusion process with known properties.

This paper proposes a strategy to estimate the intensity of default under arbitrary specifications for stationary diffusions based on the EMM, see Gallant and Tauchen (1996). The EMM estimator is particularly useful when the likelihood function cannot be written in closed form and it does not rely on known approximations of the possibly unknown likelihood function. Intuitively, this is possible because the EMM estimator uses an auxiliary model to derive general moment conditions. The auxiliary model can be estimated using semi-nonparametric methods to obtain a good statistical representation of the data without the introduction of arbitrary assumptions about its distribution. Under the correct specification of the structural model the estimator is consistent and if in addition the auxiliary model is also right the estimator is asymptotically as efficient as maximum likelihood (ML). Thus, under this latter case the broadly applied ML approach is not superior in terms of efficiency and the same comparison is true between the EMM and the Simulated Method of Moments.[7] As long as the underlying model can be used to

---

[6]In Aït-Sahalia and Kimmel (2007) the stock price and the unobserved state follow diffusion processes, and the closed form expansions for the log-likelihood function must be derived for the particular specifications of these diffusion processes. In order to map the observable option prices into the unobservable state variables, the likelihood function of the states is multiplied by a Jacobian term, which depends on the partial derivatives of the state variables with respect to the stock prices and the option prices.

[7]EMM estimator can be seen as the SMM with a different GMM criterion function. For alternative simulation-based techniques that are suitable for models with unobserved state variables, see the method of Indirect Inference in Gourieroux, Monfort, and Renault (1993), Smith (1993) and the Simulated Method of Moments in Duffie and Singleton (1993), Lee and Ingram (1991),

produce simulations, the EMM is computationally tractable, while that for complicated nonlinear models the method of Indirect Inference might be infeasible. Furthermore, the procedures to compute the estimator do not change with the specification of the reduced-form model, and then it is possible to estimate different specifications modifying the underlying reduced form model only.

The first section develops a reduced-form model for the intensity of default under the doubly stochastic framework. The following section presents a strategy to estimate the model based on the EMM. The next section discusses the empirical results and the performance of the general model under different specifications. Finally, there is a discussion on the conclusions of this study and its possible extensions.

II Reduced-form model

This section presents a reduced form model based on a broadly accepted pricing equation for credit default swaps (CDS) and the doubly stochastic framework (see for instance, Berndt, Douglas, Duffie, Ferguson, and Schranz, 2008, Longstaff, Mithal, and Neis, 2005, Pan and Singleton, 2008, Zhang, 2004). The process that describes the intensity of default[8] can be modeled as a stationary diffusion process. Denote by t a continuous counter for the time and let the unobserved arrival rate of default on sovereign bonds at time t be λ_t. The subindexes of t denote particular moments in time, for instance t_i represents the timing of the observations $i = 0, 1, \ldots$, with $t_i > t_s \iff i > s$.[9] The subindex of t is omitted when the reference to a particular moment i is not relevant.

Given some $i > 0$, an initial condition $\lambda_{t_{i-1}}$ for any $i > 0$ and some functions b_ρ and σ_ρ satisfying Assumptions A1 to A4 in Appendix A, the process for $\{\lambda_{t_i}\}_{i>0}$ is given by the solution to the time-homogeneous stochastic differential equation (SDE),

$$d \ln \lambda_t = b_\rho \left(\ln \lambda_t \right) dt + \sigma_\rho \left(\ln \lambda_t \right) dW_t, \tag{2}$$

where W_t is a standard Wiener process, the functions b_ρ, σ_ρ are the drift and diffusion "coefficients" respectively and the subindex ρ makes explicit the dependence of the unknown parameters ρ^o.[10] In principle, the process for the intensity does not need to be an Ito process and alternative processes could also be considered, but at this point this extension is left for future research.

Let M denote the maturity (in years) of a CDS contract agreed at some t and $CDS_{(M)t}$ denote the observed (annualized) premium at issue, i.e. the price that the buyer must pay for the protection of the CDS contract expressed as a rate on the notional amount to be protected. The CDS pricing equation has two components known as "legs", which under the usual arbitrage free assumption must be equal (Duffie and Singleton, 2003, Pan and Singleton, 2008). The premium leg, is the present value of the buyer's premiums to be paid contingent on the lack of a credit event, and the protection leg is the present value of the contingent payments due by the seller upon a credit event.

Pakes and Pollard (1989).

[8]Technically, defaults are not relevant credit events for sovereign CDS, the ISDA terms sheets for plain vanilla sovereign CDS contracts defines other four credits events. In this paper, these events are not differentiated and modeled together as "default" events.

[9]The observations are in discrete time, for instance daily or a higher frequency.

[10]The specification in logs prevents modeling negative hazard rates.

Consider a set of outcomes Ω, the σ–algebra $\mathcal{F}$[11] and a σ–algebra $\mathcal{F}_t$ containing the collection of events that corresponds to the information available at t, such that if $t \leq s$ then $\mathcal{F}_t \subseteq \mathcal{F}_s \subseteq \mathcal{F}$. The expectation conditional on $\mathcal{F}_t$ is denoted as $E\left[\cdot|\mathcal{F}_t\right]$, the risk free rate at t is r_t, the recovery rate is δ, and the frequency of the buyer's due payments in one year is q_c.[12] Suppose that by time t_i the market implicitly agrees on some initial condition $\ddot{\lambda}_{t_i}$ for equation (2) then equation (3) is the pricing equation of a CDS with maturity M, where the left and right hand sides are the premium and protection legs respectively. The price that equalize both legs at time t is denoted by $\overline{CDS}_t\left(M, \mathcal{F}_t, \ddot{\lambda}_t; \rho^o\right)$, where ρ^o is a vector of unknown parameters and $\ddot{\lambda}_t$ is a specified initial condition. The notation in equation (3) simplifies the notional amount to one unit of face value, see e.g. Duffie and Singleton (2003) for details about the derivation of the CDS pricing equation.[13]

$$\frac{1}{q_c}\overline{CDS}_{t_i}\left(M, \mathcal{F}_{t_i}, \ddot{\lambda}_{t_i}; \rho^o\right) \sum_{j=1}^{q_c M} E_{\rho^o}\left[e^{-\int_{t_i}^{t_i+j/q_c}(r_s+\lambda_s)ds}\middle| \mathcal{F}_{t_i}, \ddot{\lambda}_{t_i}\right] \tag{3}$$

$$= (1-\delta)\int_{t_i}^{t_i+M} E_{\rho^o}\left[\lambda_u\, e^{-\int_{t_i}^{u}(r_s+\lambda_s)ds}\middle| \mathcal{F}_{t_i}, \ddot{\lambda}_{t_i}\right]du,$$

The discount factor in equation (3) is derived from the usual discount factor in continuous time, $e^{-\int_0^t r_s ds}$,[14] and the probability of a credit event for the underlying bond within the corresponding period of time. Following the model proposed by Lando (1998), a credit event is the first jump time τ of a Cox process with arrival intensity λ_t, such that

$$\tau = \inf\left\{t: \int_{t_j}^t \lambda_s ds \geq Y_t\right\},$$

and Y_t is an unit exponential random variable.[15] Then for a given $\{\lambda_t\}_{t=t_j}^{t_J}$ and a particular distribution of Y_t (with its memoryless property), the survival probability between time t_j and t_J is

$$\Pr\left\{\tau \notin (t_j, t_J]\middle| \{\lambda_t\}_{t=t_j}^{t_J}\right\} = \Pr\left\{Y_t > \int_{t_j}^t \lambda_s ds, \forall t \in (t_j, t_J]\middle| \{\lambda_t\}_{t=t_j}^{t_J}\right\}$$

$$= e^{-\int_{t_j}^{t_J}\lambda_s ds}.$$

In the doubly stochastic framework, λ_t is also a random variable, so the probability of avoiding a credit event between time t_j and t_J becomes

$$\underbrace{\Pr\left\{\tau \notin (t_j, t_J]\middle| \mathcal{F}_{t_i}, \lambda_{t_i}\right\} = E\left[e^{-\int_{t_j}^{t_J}\lambda_s ds}\middle| \mathcal{F}_{t_i}, \lambda_{t_i}\right]}_{survival\ probability}.$$

---

[11] A collection of subsets of Ω that can be assigned a probability.

[12] The recovery rate δ is the rate at which the value of the underlying bond is assumed to be recovered in case of the credit event. This is a rate that appears in the CDS agreements.

[13] Under the doubly stochastic framework equation (3) can be derived from the more intuitive formulation, $\frac{1}{q_c}\overline{CDS}_{t_i}\left(M, \mathcal{F}_{t_i}, \ddot{\lambda}_{t_i}; \rho^o\right)\sum_{j=1}^{q_c M} E_{\rho^o}\left[e^{-\int_{t_i}^{t_i+j/q_c} r_s ds}\mathbb{I}\{\tau > t_i + j/q_c\}\middle| \mathcal{F}_{t_i}, \ddot{\lambda}_{t_i}\right] = E_{\rho^o}\left[e^{-\int_{t_i}^{\tau} r_s ds}(1-\delta)\mathbb{I}\{\tau \leq M\}\middle| \mathcal{F}_{t_i}, \ddot{\lambda}_{t_i}\right]$, where $\mathbb{I}$ is the indicator function.

[14] If a value x_{t-1} is invested in a risk free asset, then $x_t = x_{t-1}(1 + r_{t-1})$ in discrete time, and $\frac{d\ln x_t}{dt} = r_t$ in continuous time.

[15] A Cox process is a generalization of a Poisson process with random intensity such that conditioning on a realization of the intensity the jump process becomes an inhomogeneous Poisson process with intensity equal to the conditioning intensity.

Assume that the risk-free rate r_t and λ_t are independent processes. This assumption is not strong considering that for any point in time the information of the most up-to-date forward risk free interest rates can be used as the relevant rates to compute the CDS prices. Following Pan and Singleton (2008), equation (3) can be simplified in terms of the prices of a default-free zero-coupon bond issued at t_i with maturity at t_k, denoted by $\bar{p}\left(t_i, t_k\right)$.

$$\overline{CDS}_{t_i}\left(M, \mathcal{F}_{t_i}, \ddot{\lambda}_{t_i}; \rho^o\right) = \frac{(1-\delta) \int_{t_i}^{t_i+M} \bar{p}\left(t_i, u\right) E_{\rho^o}\left[\lambda_u \, e^{-\int_{t_i}^{u} \lambda_s ds} \middle| \mathcal{F}_{t_i}, \ddot{\lambda}_{t_i}\right] du}{\frac{1}{q_c} \sum_{j=1}^{q_c M} \bar{p}\left(t_i, t_j\right) E_{\rho^o}\left[e^{-\int_{t_i}^{t_j} \lambda_s ds} \middle| \mathcal{F}_{t_i}, \ddot{\lambda}_{t_i}\right]} \tag{4}$$

with $p\left(t_i, t_k; \boldsymbol{r}_{t_i}\right) = e^{-\int_{t_i}^{t_k} r_{s,t_i} ds}$, $\bar{p}\left(t_i, t_k\right) = E\left[p\left(t_i, t_k; \boldsymbol{r}_{t_i}\right) \middle| \mathcal{F}_{t_i}\right]$, $\boldsymbol{r}_{t_i}$ as the risk free zero-coupon yield curve and r_{s,t_i} as the spot rate at maturity s of the curve $\boldsymbol{r}_{t_i}$. By time t_i the curves of forward risk free interest rates for $t \leq t_i$ are known and if the corresponding curve is actually used by the agents whenever $\boldsymbol{r}_{t_i}$ is needed, a reasonable assumption is that $\bar{p}\left(t_i, t_k\right) = p\left(t_i, t_k; \boldsymbol{r}_{t_i}\right)$.[16]

The main building blocks of the reduced form model are completed by specifying the object that reveals the initial conditions for equation (2). The problem of obtaining the initial condition λ_{t_0} can be solved using the information in the price of the corresponding underlying bond of the CDS contracts. The intensity of default also has an important role in the pricing of risky bonds and these prices can be modeled with similar tools. Following Jarrow (2004), the pricing formula for an underlying zero-coupon risky bond is given by equation (5)

$$\begin{aligned} \overline{B}_{t_i}\left(M, \mathcal{F}_{t_i}, \lambda_{t_i}; \rho^o\right) = \\ E_{\rho^o}\left[e^{-\int_{t_i}^{t_i+M} r_s ds} \mathbb{I}\{\tau > t_i + M\} + \delta \, e^{-\int_{t_i}^{\tau} r_s ds} \mathbb{I}\{\tau \leq t_i + M\} \middle| \mathcal{F}_{t_i}, \lambda_{t_i}\right], \end{aligned} \tag{5}$$

where the indicator function is denoted by $\mathbb{I}$. In practice the underlying bonds of the CDS usually pay coupons, in such case equation (5) must be generalized accordingly, but to keep the presentation of the model simpler these details are not considered at this point.[17]

Assume that the reference price at t_{i-1} for the underlying bond with maturity M, i.e. $B_{(M)t_{i-1}}$, is known with certainty at t_i. Under the doubly stochastic framework and assuming that $\boldsymbol{r}_t$ and λ_t are independent processes, the price of the bond at t_{i-1} becomes

$$\begin{aligned} \overline{B}_{t_{i-1}}\left(M, \mathcal{F}_{t_i}; \rho^o\right) \equiv p\left(t_{i-1}, t_{i-1} + M; \boldsymbol{r}_{t_{i-1}}\right) E_{\rho^o}\left[e^{-\int_{t_{i-1}}^{t_{i-1}+M} \lambda_s \, ds} \middle| \mathcal{F}_{t_i}\right] \\ + \delta \int_{t_{i-1}}^{t_{i-1}+M} p\left(t_{i-1}, u; \boldsymbol{r}_{t_{i-1}}\right) E_{\rho^o}\left[\lambda_u \, e^{-\int_{t_{i-1}}^{u} \lambda_s \, ds} \middle| \mathcal{F}_{t_i}\right] du. \end{aligned} \tag{6}$$

The solution of the expectation in equation (6) depends on the initial condition $\lambda_{t_{i-1}}$, this condition does not appear explicitly because it is considered an object in $\mathcal{F}_{t_i}$. The right hand side of equation (6) can be

---

[16]It is not the goal of this reduced form approach to estimate a stochastic model in continuous time for the term structure and the literature suggest that increasing the complexity of the model to include this feature might not be relevant. Using a the two-factor affine model, Pan and Singleton (2008) checked the sensitivity of their one-factor model's results to the presence of stochastic interest rates and they conclude "... we obtained virtually identical results to those with a constant risk-free rate."

[17]The application in this paper uses bonds that pay coupons and applies a generalization of equation (5). For further details on pricing equations for risky bonds that pay coupons see Jarrow (2004).

solved for a given ρ and λ_{t_i}. Provided that equation (7) holds exactly, the problem of determining λ_{t_0} for a given ρ is reduced to the inversion of equation (6) for the underlying risky bond. For $i = 1$,

$$B_{(M)t_{i-1}} = \overline{B}_{t_{i-1}}\left(M, \mathcal{F}_{t_i}; \rho\right). \tag{7}$$

Therefore, by time t_1, if ρ^o is given then the corresponding λ_{t_0} that solves equation (7) is known.

To complete the setup, assume that the market knows the true ρ^o and that every period t_i there is an implicit agreement about some $\ddot{\lambda}_{t_i}$, which is drawn from the conditional distribution implied by the process in equation (2) under the right initial condition $\lambda_{t_{i-1}}$. This simplification makes sense because the reference price $B_{(M)t_{i-1}}$ is assumed not observed until t_i and therefore the $\lambda_{t_{i-1}}$ of reference is unknown until t_i.[18]

In the case of the sovereign bonds, most counties do not have many, j, observations of the CDS prices per date, then one can think about some $\ddot{\lambda}_{t_i}$ driving these prices and some additive noise, $\epsilon_{t_i,j}$, with 0 conditional mean that explains the differences between the observed quotes, $CDS_{(M)t_i,j}$.

$$CDS_{(M)t,j} = \overline{CDS}_t\left(M, \mathcal{F}_t, \ddot{\lambda}_t; \rho^o\right) + \epsilon_{t,j} \tag{8}$$

Thus in the application the observed CDS price of reference for t_i is denoted by $CDS_{(M)t_i}$ and corresponds to the average of the observed quotes $CDS_{(M)t_i,j}$ for all j.

III Estimation strategy

The construction of the estimator of this model faces two main complications. The first difficulty arises because for some given functions b_ρ and σ_ρ the transition density of equation (2) is in general unknown in closed form. Moreover, the diffusion process λ_t is an unobservable latent process, and the attempts to estimate the parameters in b_ρ and σ_ρ require a model that maps the outcomes of the intensity process to the observables. The available data is $\{CDS_{(M)t_i}, B_{(M)t_i}, \boldsymbol{r}_{t_i}\}_{i=0}^I$ and this section presents an estimation method under which the Efficient Method of Moments (EMM, Gallant and Tauchen 1996) can solve this complexity.

The first step towards the application of the EMM to this problem is to derive a density that has an analytic closed form and is a good approximation of the density of the data. This density provides an auxiliary model to construct the moment conditions of the EMM estimator for the parameters of the reduced form model, ρ^o.

The density of the observables is unknown in closed form and there is not an evident choice for an auxiliary model to model such density, thus firstly an auxiliary model is estimated through a semi-nonparametric (SNP) method. Let $\tilde{y}_t$ represent the vector of observations $(\ln CDS_{(M)t}, \ln B_{(M)t})'$ and L be a number of lags,[19] with $\tilde{\mathbf{y}}_{t_{k-1}^L} = (\tilde{y}_{t_{k-1}}, \tilde{y}_{t_{k-2}}, \ldots, \tilde{y}_{t_{k-l}}, \ldots, \tilde{y}_{t_{k-L}})'$, $k \geq L$ and $\mathcal{T} = I + 1 - L$. When is useful to emphasize the difference between observations of a random variable X and their

---

[18]Allowing for $\ddot{\lambda}_{t_i} \neq \lambda_{t_i}$ opens different possibilities for modeling how $\ddot{\lambda}_{t_i}$ is determined but within the proposed setup this distinction does not seem important for daily observations and its relevance should decrease with higher frequencies.

[19]The number L is determined statistically by the Schwarz Bayes information criterion Schwarz (1978).

simulations, a $\tilde{x}$ points out the former ones and a $\check{x}$ the latter ones. Assume that the process $\tilde{y}_t$ is strictly stationary, then the law of motion is the one-step ahead conditional density given its past. Represent this time invariant law as a suitable series expansion of some degree, which might have an infinite number of terms, and denote it by $f(\cdot|\cdot;\theta^o)$, with θ^o as the possibly infinite-dimensional vector of unknown true parameters that define its functional form. SNP estimators are constructed using series expansions that can consistently estimate density functions (see section IV.1 for a brief explanation and Gallant and Nychka, 1987, Gallant and Tauchen, 1988 for the theoretical foundations). In this application the series are Hermite polynomials. Define θ^o as

$$\theta^o = \arg\max_{\theta \in \Theta} \int \int \ln f\left(y_{t_i}|\mathbf{y}_{t_{i-1}^L};\theta\right) d\varrho_y\left(\mathbf{y}_{t_{i-1}^L}, y_{t_i}\right) \tag{9}$$

where ϱ_y denotes the true unknown joint distribution of $\tilde{y}_t$.[20] Equation (10) displays the quasimaximum likelihood estimates of θ^o, which are denoted by $\hat{\theta}_{\mathcal{T}}$.

$$\hat{\theta}_{\mathcal{T}} = \arg\max_{\theta \in \Theta} \frac{1}{\mathcal{T}} \sum_{i=L}^{I} \ln f\left(\tilde{y}_{t_i}|\tilde{\mathbf{y}}_{t_{i-1}^L};\theta\right), \tag{10}$$

The SNP estimator estimates consistently θ^o under regular conditions (see Gallant and Nychka, 1987) and its scores in equation (11) are useful to construct the moment conditions of the EMM. Provided a sample size that allows to include enough terms in the expansions of f, the caveats regarding the ability of an auxiliary model $f(\cdot|\cdot;\hat{\theta}_{\mathcal{T}})$ to approximate the intended distribution are minimal (Gallant and Long 1997).

$$\frac{1}{\mathcal{T}} \sum_{i=L}^{I} (\partial/\partial\theta) \ln f\left(\tilde{y}_{t_i}|\tilde{\mathbf{y}}_{t_{i-1}^L};\hat{\theta}_{\mathcal{T}}\right) \tag{11}$$

Provided the first order conditions implied by the optimization problem in equation (10) and an auxiliary model f that is a good statistical description of the data, the scores evaluated at $\hat{\theta}_{\mathcal{T}}$, i.e. equation (11), should be as close to $\mathbf{0}$ as the sample size allows it. The moment conditions for the EMM estimator consist of the average of the score but over the simulations $\check{y}_t$ produced by a structural or reduced form model instead of over the data $\tilde{y}_t$. Let the joint probability distribution of $(\mathbf{y}_{t_{i-1}^L}, y_{t_i})$ implied by the reduced form model proposed in section II be $q_y(\cdot,\cdot;\rho)$, then if $\rho = \rho^o$ and the model is correct, i.e. $dq_y(\cdot,\cdot;\rho^o) = d\varrho_y(\cdot,\cdot)$, there should not be much difference between the distribution of the data and that corresponding to its simulations. Thus the moment conditions defined with the scores evaluated at the simulations should be also close to $\mathbf{0}$.

Assume that the reduced form model is correct, then $m_{\mathcal{T}}(\rho, \hat{\theta}_{\mathcal{T}})$ defined in equation (12) defines moment functions for a Markov process y_t of order L and conditional density function q_y

$$m_{\mathcal{T}}\left(\rho, \hat{\theta}_{\mathcal{T}}\right) =$$
$$\frac{1}{\mathcal{T}} \sum_{i=L}^{I} \int \cdots \int (\partial/\partial\theta) \ln f\left(y_{t_i}|\mathbf{y}_{t_{i-1}^L};\hat{\theta}_{\mathcal{T}}\right) q_y\left(y_{t_i}|\mathbf{y}_{t_{i-1}^L};\rho\right) q_y\left(y_{t_{i-1}}|\mathbf{y}_{t_{i-2}^L};\rho\right) \tag{12}$$
$$\cdots q_y\left(y_{t_L}|\mathbf{y}_{t_{L-1}^L};\rho\right) q_0\left(\mathbf{y}_{t_{L-1}^L};\rho\right) d\left(y_{t_i} \cdots y_{t_0}\right),$$

---

[20]The treatment of the initial conditions is disregarded here to simplify the exposition of the main points.

where the distribution of the initial conditions is given by q_0 and can be a degenerate distribution. Note that the different treatments of the initial conditions are negligible as the sample size $\mathcal{T}$ grows. The consistency of the EMM depends on the validity of its moment conditions and under the true parameter values for a correct reduced form model $m_{\mathcal{T}}(\rho^o, \hat{\theta}_{\mathcal{T}}) \to 0$ as $\mathcal{T} \to \infty$. A key identification assumption requires that the number of parameters in the vector θ is larger than the number of parameters in ρ, such that there is a sufficient number of moment functions.[21]

The transition q_y is unknown in closed form and the moment condition in equation (12) can be approximated with a negligible error by Monte Carlo (or alternative methods). Let $\{\{\check{y}_{t_i,k}\}_{i=0}^{I}\}_{k=1}^{K}$ be K sets of simulations of length $\mathcal{T}$ from $q_y(\cdot|\cdot;\rho)$ and $q_0(\cdot|\cdot;\rho)$, then if K is large enough[22] equation (13) holds.

$$m_{\mathcal{T}}\left(\rho, \hat{\theta}_{\mathcal{T}}\right) = \frac{1}{\mathcal{T}} \sum_{i=L}^{I} \frac{1}{K} \sum_{k=1}^{K} (\partial/\partial\theta) \ln f\left(\check{y}_{t_i,k} | \check{\mathbf{y}}_{t_{i-1,k}^L}; \hat{\theta}_{\mathcal{T}}\right) \tag{13}$$

The reduced form model proposed in section II does not fit the basic setup for the application of the EMM estimator because it does not model the law of motion of $\mathbf{r}_t$. The EMM method can handle exogenous variables as proposed in Gallant and Tauchen (1996), but the estimation of a SNP density with a large number of variables is unfeasible and therefore the definition of conditional moment conditions based on SNP estimates finds limitations. One solution to this issue comes from noticing that the reduced form model defines a mapping Γ_ρ such that $y_{t_i} = \Gamma_\rho(\mathbf{r}_{t_i}, \boldsymbol{\lambda}_{t_i})$ for $\boldsymbol{\lambda}_{t_i} = (\lambda_{t_i}, \ddot{\lambda}_{t_i})'$, then equation (12) still can be computed after a suitable change of random variables and their corresponding measures. The theoretical model assumes that the risk free forward curves of reference are observable, independent from the intensity and its distribution is exogenously given. As a result, the proposed auxiliary model does not need to explicitly consider the curves $\mathbf{r}_t$. Note that $\Gamma_\rho(\cdot) = (\ln \overline{CDS}_{t_i}(\cdot;\rho), \ln \overline{B}_{t_i}(\cdot;\rho))'$ and denote its L lags by $\Gamma_{\rho_-}(\mathbf{r}_{t_{i-k}^L}, \boldsymbol{\lambda}_{t_{i-k}^L})$,[23] such that $\mathbf{y}_{t_{i-k}^L} = \Gamma_{\rho_-}(\mathbf{r}_{t_{i-k}^L}, \boldsymbol{\lambda}_{t_{i-k}^L})$. Let $q_\lambda(\cdot|\cdot;\rho)$ and $q_{\lambda_0}(\cdot;\rho)$ be the conditional density for the intensity and the density of $\boldsymbol{\lambda}_{t_0}$ respectively. Similarly, $\varrho_r(\cdot,\cdot)$, $\varrho_{r_0}(\cdot)$ and $\varrho_r(\cdot|\cdot)$ denote the true unknown joint distribution, density and conditional density of the forward curves respectively. Recall that the model assumes that $\mathbf{r}_t$ is independent from λ_t, then equation (12) can be rewritten as

$$m_{\mathcal{T}}\left(\rho, \hat{\theta}_{\mathcal{T}}\right) = \frac{1}{\mathcal{T}} \sum_{i=L}^{I} \int \cdots \int (\partial/\partial\theta) \ln f\left(\Gamma_\rho(\mathbf{r}_{t_i}, \boldsymbol{\lambda}_{t_i}) | \Gamma_{\rho_-}(\mathbf{r}_{t_{i-1}^L}, \boldsymbol{\lambda}_{t_{i-1}^L}); \hat{\theta}_{\mathcal{T}}\right) \tag{14}$$
$$\left(\prod_{j=1}^{i} q_\lambda\left(\boldsymbol{\lambda}_{t_j} | \lambda_{t_{j-1}}; \rho\right) d\boldsymbol{\lambda}_{t_j}\right) q_{\lambda_0}\left(\boldsymbol{\lambda}_{t_0}; \rho\right) d\boldsymbol{\lambda}_{t_0} d\boldsymbol{\varrho}_r(\mathbf{r}_{t_{i-1}^L}, \mathbf{r}_{t_i}),$$

then put $\check{y}_{t_i,k} = \Gamma_\rho(\mathbf{r}_{t_i}, \check{\boldsymbol{\lambda}}_{t_i,k})$ and equation (13) still holds to compute the intended moment conditions. Moreover, equation (14) can be averaged over a panel of simulations to reduce variance.

---

[21] Note that 1 degree of freedom is lost because the SNP density is constructed with normalized Hermite polynomials.

[22] K should be large enough to ensure that the difference between the Monte Carlo integral (or the numerical integral of any alternative method) and the analytic integral is negligible.

[23] $\Gamma_{\rho_-}\left(\mathbf{r}_{t_{i-k}^L}, \boldsymbol{\lambda}_{t_{i-k}^L}\right) = (\Gamma_\rho(\mathbf{r}_{t_{i-k}}, \boldsymbol{\lambda}_{t_{i-k}})', \ldots, \Gamma_\rho(\mathbf{r}_{t_{i-k-L+1}}, \boldsymbol{\lambda}_{t_{i-k-L+1}})')'.$

To summarize the procedure, the reduced form model is compactly rewritten in equations (15) to (18).

$$\check{y}_{t_i,k} = \Gamma_\rho(\boldsymbol{r}_{t_i}, \check{\boldsymbol{\lambda}}_{t_i,k}) \tag{15}$$

$$\ln \check{\lambda}_{t_i,k} = \ln \check{\lambda}_{t_{i-1},k} + \int_{t_{i-1}}^{t_i} b_\rho\left(\ln \lambda_t\right) dt + \int_{t_{i-1}}^{t_i} \sigma_\rho\left(\ln \lambda_t\right) d\ddot{W}_{t,k} \tag{16}$$

$$\ln \check{\lambda}_{t_i,k} = \ln \check{\lambda}_{t_{i-1},k} + \int_{t_{i-1}}^{t_i} b_\rho\left(\ln \lambda_t\right) dt + \int_{t_{i-1}}^{t_i} \sigma_\rho\left(\ln \lambda_t\right) dW_{t,k} \tag{17}$$

$$\check{\lambda}_{t_0,k} = \overline{B}_{t_0}^{-1}\left(M, B_{(M)t_0}, \boldsymbol{r}_{t_0}; \rho\right). \tag{18}$$

To draw simulations from the reduced form model, the first step is to compute an initial condition $\check{\lambda}_{t_0}$ as in equation (18), i.e. inverting the pricing equation for the bonds for a given ρ, $\boldsymbol{r}_{t_0}$ and $B_{(M)t_0}$. Then pick a suitable $\check{\lambda}_{t_0}$ such that $\check{y}_{t_0} = \check{y}_{t_0}$.[24] Provided a $\check{\lambda}_{t_{i-1}}$, draw $\check{\boldsymbol{\lambda}}_{t_i}$ from equations (16) and (17) for all $i \geq 1$. Using the simulations $\check{\boldsymbol{\lambda}}_{t_i}$ compute $\check{y}_{t_i}$ according to Γ_ρ, i.e. compute $\overline{CDS}_{t_i}\left(M, (\boldsymbol{r}'_{t_i}, \check{\lambda}_{t_{i-1}}, \cdot)', \check{\lambda}_{t_i}; \rho\right)$ and $\overline{B}_{t_i}\left(M, (\boldsymbol{r}'_{t_i}, \cdot)', \check{\lambda}_{t_i}; \rho\right)$ as defined in equations (4) and (5).

The expectations in the pricing formulas, equations (4) and (5), have a closed form solution only under some particular specifications for the diffusion that drives the intensity. The proposed method works with any stationary diffusion. One of the computational issues for drawing simulations from the reduced form model is to solve the expectations in the pricing formulas without knowing their closed form solutions. Monte Carlo simulations were not be the best choice to compute these expectations. Here they are solved by the Crank-Nicolson implicit finite-difference method and the Feynman-Kac results.

The consistency of the EMM depends on the validity of its moment conditions,[25] and the auxiliary model f does not need to nest the distribution implied by the reduced form model q_y to obtain a consistent estimator. Intuitively, it is not necessary to choose a correct auxiliary model $f\left(\cdot|\cdot; \theta_{\mathcal{T}}^o\right)$, since the moment functions in equation (13) mimic the first-order conditions of the optimization problem in equation (10) and they must hold if the reduced form model is the true one. Then a consistent estimator of ρ^o is still available by choosing $\hat{\rho}_{\mathcal{T}}$ such that $m_{\mathcal{T}}(\hat{\rho}_{\mathcal{T}}, \hat{\theta}_{\mathcal{T}}) \simeq \boldsymbol{0}$.

If further assumptions hold, the estimator becomes as efficient as maximum likelihood. If there is a smooth mapping[26] of the reduced form parameters into the parameters of the auxiliary model $f\left(\cdot|\cdot; \theta_{\mathcal{T}}^o\right)$, Gallant and Tauchen (1996) prove that the EMM and the ML estimators have the same asymptotic distribution. Hence, the closer is $f\left(\cdot; \theta^o\right)$ to the assumed true data generating process, $q_y\left(\cdot; \rho^o\right)$, the higher is the efficiency of the EMM estimator. In this case, a fully efficient estimator is achieved if $\left\{\varrho_{r_0}(\boldsymbol{r}_{t_{L-1}^L}) q_{\lambda_0}\left(\boldsymbol{\lambda}_{t_0}; \rho\right) \prod_{j=1}^{L-1} q_\lambda\left(\boldsymbol{\lambda}_{t_j}|\lambda_{t_{j-1}}; \rho\right), \left\{q_\lambda\left(\boldsymbol{\lambda}_{t_j}|\lambda_{t_{j-1}}; \rho\right) \varrho_r(\boldsymbol{r}_{t_j}|\boldsymbol{r}_{t_{j-1}^L})\right\}_{j=L}^\infty\right\}_{\rho \in \mathbb{R}^{d_\rho}}$ is smoothly embedded within the auxiliary model $\{f_0(\Gamma_{\rho_-}(\boldsymbol{r}_{t_{L-1}^L}, \boldsymbol{\lambda}_{t_{L-1}^L})), \{f(\check{\Gamma}_\rho(\boldsymbol{r}_{t_i}, \boldsymbol{\lambda}_{t_i})|\Gamma_{\rho_-}(\boldsymbol{r}_{t_{i-1}^L}, \boldsymbol{\lambda}_{t_{i-1}^L}); \theta)\}_{i=L}^\infty\}_{\theta \in \Theta}$.

The asymptotic distribution of $\hat{\theta}_{\mathcal{T}}$ in equation (19) is a key result to complete the construction of the EMM estimator that uses all the moment equations in equation (13) efficiently. Let $\mathcal{H}_{\mathcal{T}}^o = \partial m_{\mathcal{T}}\left(\rho^o, \theta_{\mathcal{T}}^o\right)/\partial\theta'$

---

[24]Note that $\check{\lambda}_{t_0}$ can be fixed to a value such that equation (3) holds for $CDS_{(M)t_0} = \overline{CDS}_{t_0}(M, \mathcal{F}_{t_0}, \check{\lambda}_{t_0}; \rho)$. There are not random variables in the reduced form model that depend on previous realizations of $\check{\lambda}_t$.

[25]i.e. the existence of a sequence $\{\theta_{\mathcal{T}}^o\}$ such that $m_{\mathcal{T}}\left(\rho, \theta_{\mathcal{T}}^o\right) = 0$ implies that $\rho = \rho^o$ for all $\mathcal{T}$ larger that some finite amount, and $\lim_{\mathcal{T}\to\infty}(\hat{\theta}_{\mathcal{T}} - \theta_{\mathcal{T}}^o) = 0$ almost surely. See the regularity conditions for this result in Gallant and Tauchen (1996).

[26]A smooth mapping in the sense of definition 1, Smoothly Embedded, in Gallant and Tauchen (1996).

and

$$\mathcal{J}_{\mathcal{T}}^{o} = Var \left[\frac{1}{\sqrt{\mathcal{T}}} \sum_{i=L}^{I} \frac{\partial \left(\ln f \left(\tilde{y}_{t_i} | \tilde{\mathbf{y}}_{t_{i-1}^L}, \theta_{\mathcal{T}}^{o} \right) \right)}{\partial \theta} \right]$$

thus it can be shown that

$$\sqrt{\mathcal{T}} \left(\hat{\theta}_{\mathcal{T}} - \theta_{\mathcal{T}}^{o} \right) \xrightarrow{d} \mathbf{N} \left[0, \left(\mathcal{H}_{\mathcal{T}}^{o} \right)^{-1} \mathcal{J}_{\mathcal{T}}^{o} \left(\mathcal{H}_{\mathcal{T}}^{o} \right)^{-1} \right]. \tag{19}$$

By Taylor's theorem and $m_{\mathcal{T}} \left(\rho^{o}, \theta_{\mathcal{T}}^{o} \right) = 0$ it follows that

$$\sqrt{\mathcal{T}} m_{\mathcal{T}} \left(\rho^{o}, \hat{\theta}_{\mathcal{T}} \right) = \sqrt{\mathcal{T}} m_{\mathcal{T}} \left(\rho^{o}, \theta_{\mathcal{T}}^{o} \right) + \left[\mathcal{H}_{\mathcal{T}}^{o} + o_p \left(1 \right) \right] \sqrt{\mathcal{T}} \left(\hat{\theta}_{\mathcal{T}} - \theta_{\mathcal{T}}^{o} \right)$$
$$= \mathcal{H}_{\mathcal{T}}^{o} \sqrt{\mathcal{T}} \left(\hat{\theta}_{\mathcal{T}} - \theta_{\mathcal{T}}^{o} \right) + o_p \left(1 \right),$$

which by equation (19) implies

$$\sqrt{\mathcal{T}} m_{\mathcal{T}} \left(\rho^{o}, \hat{\theta}_{\mathcal{T}} \right) \overset{d}{\approx} \mathbf{N} \left[0, \mathcal{J}_{\mathcal{T}}^{o} \right].$$

Therefore, given a consistent estimator $\mathcal{J}_{\mathcal{T}}$ of $\mathcal{J}_{\mathcal{T}}^{o}$, the estimator of the EMM is

$$\hat{\rho}_{\mathcal{T}} = \arg\min_{\rho \in \Xi} m'_{\mathcal{T}} \left(\rho, \hat{\theta}_{\mathcal{T}} \right) \mathcal{J}_{\mathcal{T}}^{-1} m_{\mathcal{T}} \left(\rho, \hat{\theta}_{\mathcal{T}} \right), \tag{20}$$

where[27]

$$\mathcal{J}_{\mathcal{T}} = \frac{1}{\mathcal{T}} \sum_{i=L}^{I} \left[(\partial/\partial\theta) \ln f \left(\tilde{y}_{t_i} | \tilde{\mathbf{y}}_{t_{i-1}^L}, \hat{\theta}_{\mathcal{T}} \right) \right] \left[(\partial/\partial\theta) \ln f \left(\tilde{y}_{t_i} | \tilde{\mathbf{y}}_{t_{i-1}^L}, \hat{\theta}_{\mathcal{T}} \right) \right]'.$$

The goal is to obtain estimates of the reduced form parameters $\hat{\rho}_{\mathcal{T}}$ for alternative specifications of b_{ρ} and σ_{ρ} in equation (2). Once that there is a $\hat{\rho}_{\mathcal{T}}$ and some b_{ρ} and σ_{ρ} which define a stationary process (see assumptions A1 to A4 in Appendix A), then the corresponding estimate for the law of motion of the intensity is identified. The main advantage of using the EMM is that the estimator works for any ρ° that defines a stationary process even when the corresponding law of motion is allowed to be unknown. The complexity of applying the estimator is reduced by the EMM package described in Gallant and Tauchen (2013a) which is written in C++ and implements the algorithm in Chernozhukov and Hong (2003) to solve the optimization problem in equation (20).[28] The key point to implement the proposed estimation strategy is the possibility of drawing simulations $\{\{\tilde{y}_{t_i,k}\}_{i=0}^{I}\}_{k=1}^{K}$ for each candidate ρ from the proposed reduced form model. The model in this paper and the algorithms for drawing these simulations have been programed in C++.

---

[27]By Theorem 1 in Gallant and Tauchen (1996),

$$\sqrt{\mathcal{T}} \left(\hat{\rho}_{\mathcal{T}} - \rho^{\circ} \right) \xrightarrow{d} \mathbf{N} \left(0, \left[\left(M_{\mathcal{T}}^{o} \right)' \left(\mathcal{J}_{\mathcal{T}}^{o} \right)^{-1} \left(M_{\mathcal{T}}^{o} \right) \right]^{-1} \right),$$

where $M_{\mathcal{T}} \left(\rho, \theta \right) = (\partial/\partial\rho') m_{\mathcal{T}} \left(\rho, \theta \right)$ and $M_{\mathcal{T}}^{o} = M_{\mathcal{T}} (\rho^{o}, \theta_{\mathcal{T}}^{o})$. Besides, for $\hat{M}_{\mathcal{T}} = M_{\mathcal{T}} (\hat{\rho}_{\mathcal{T}}, \hat{\theta}_{\mathcal{T}})$, $\lim_{\mathcal{T} \to \infty} (\hat{M}_{\mathcal{T}} - M_{\mathcal{T}}^{o}) = 0$ a.s..

[28]The programs used to compute the estimates for the SNP auxiliary model, $\hat{\theta}_{\mathcal{T}}$, are also borrowed from the authors of the EMM package (see Gallant and Tauchen 2013b).

IV Empirical results

IV.1 SNP estimation of the auxiliary model

The conditional density of $\tilde{y}_t$ is estimated through a SNP density modeled with orthonormal Hermite polinomials. For a brief reference of this method, let y be the M–dimensional vector $\tilde{y}_{t_i}$ for any particular i and denote by x a $M \times L$–matrix of exogenous or predetermined variables. Let $\mu_{y|x}$ and $\sum_{y|x}$ be the conditional mean and covariance matrix of y respectively and z be the corresponding vector of normalized variables.

$$z = R_x^{-1}(y - \mu_{y|x}),$$

with upper triangular R_x and $\sum_{y|x} = R_x R_x'$. Let $K = (K_z, K_x)$ and denote a multivariate polynomial in z of degree K_z whose coefficients a_x are polynomials of degree K_x in x by $\mathcal{P}_K(z, a_x)$.[29] Let $\phi(z)$ be the density function of the (multivariate) Gaussian distribution with mean zero and the identity as its variance-covariance matrix, then an Hermite function has the form $\mathcal{P}_K(z, a_x)\sqrt{\phi(z)}$. The Hermite conditional density is

$$h_K(z | a_x) = \frac{[\mathcal{P}_K(z, a_x)]^2 \phi(z)}{\int [\mathcal{P}_K(s, a_x)]^2 \phi(s) ds}, \tag{21}$$

and under standard regularity conditions (see Fenton and Gallant, 1996, Gallant and Nychka, 1987), it can be proved that there is a density function $h(z|x)$ such that

$$h(z|x) = \lim_{K_z \to \infty, K_x \to \infty} h_K(z | a_x). \tag{22}$$

Equation (22) holds because the Hermite functions are dense in the $\mathcal{L}^2$–space. In practice, K_z and K_x are finite numbers determined during the estimation procedure, so for each K there is a corresponding number of parameters $\theta_{\mathcal{T}}$ to be estimated. According to equation (21), the density function $f(\cdot|\theta_{\mathcal{T}})$ to construct the estimator is explicitly

$$f(y|x, \theta_{\mathcal{T}}) = \frac{\left\{\mathcal{P}_K\left(R_x^{-1}(y - \mu_{y|x}), a_x\right)\right\}^2 \phi\left(R_x^{-1}\left(y - \mu_{y|x}\right)\right)}{|\det(R_x)| \int \left\{\mathcal{P}_K(s, a_x)\right\}^2 \phi(s) ds}. \tag{23}$$

with estimates $\theta_{\mathcal{T}} = \hat{\theta}_{\mathcal{T}}$ computed according to equation (10). The matrix of coefficients $a_x = [a_{(\cdot)(\mathbf{0})}|A]$ has typical element $a_{(\eta)(\kappa)}$, with $(\eta) = (\eta_1, \eta_2, \ldots, \eta_M)$ and $(\kappa) = (\kappa_1, \kappa_2, \ldots, \kappa_{Lp})$, such that if $[\kappa] = \sum_{i=1}^{Lp} \kappa_i$ then $a_{(\cdot)(\mathbf{0})}$ and A contain all coefficients $a_{(\eta)(\kappa)}$ with $[\kappa] = 0$ and $[\kappa] > 0$ respectively. The degree of the polynomials in a_x is K_x with $a_{x(0,\ldots,0)(0,\ldots,0)} = 1$ as a normalization rule. For instance, if the degree of the polynomial is 0, i.e. $K_z = 0$, then by equation (23), $y|x \sim N\left(\mu_{y|x}, R_x R_x'\right)$ and equation (10) estimates a Gaussian VAR, Gaussian ARCH or a Gaussian GARCH as a Maximum Likelihood estimator would do it.

The parameters of the location function $\mu_{y|x}$ are denoted by the vector $\mathfrak{b}_0$ and the matrix B of dimensions $M \times 1$ and $M \times L_u$ respectively, such that

$$\mu_{y|x} = \mathfrak{b}_0 + Bx.$$

---

[29]Explicitly, P_K is a normalized polynomial such that $P_K(z, a_x) = \sum_{[\eta]=0}^{K_z} \alpha_{x(\eta)} z^{(\eta)}$, with $z^{(\eta)} = (z_1)^{\eta_1}(z_2)^{\eta_2}\cdots(z_M)^{\eta_M}$, $[\eta] = \sum_{i=1}^M \eta_i$, and $\alpha_{x(0,\ldots,0)} = 1$. Similarly, for all (η), $\alpha_{x(\eta)} = \sum_{[\kappa]=0}^{K_x} \alpha_{(\eta)(\kappa)} x^{(\kappa)}$.

The variance matrix $\sum_{y|x}$ has a modified BEKK expression described in Gallant and Tauchen (2013b) based on Engle and Kroner (1995), which allows for leverage and level effects.

$$\sum\nolimits_{y|x_{t_i}} = R_0 R_0^{'} \tag{24}$$

$$+ \sum_{l=1}^{L_g} Q_l \sum\nolimits_{y|x_{t_{i-l}}} Q_l^{'}$$

$$+ \sum_{l=1}^{L_r} P_l \left[y_{t_{i-l}} - \mu_{y|x_{t_{i-l}}} \right] \left[y_{t_{i-l}} - \mu_{y|x_{t_{i-l}}} \right]^{'} P_l^{'}$$

$$+ \sum_{l=1}^{L_v} \max\left\{0, V_l \left[y_{t_{i-l}} - \mu_{y|x_{t_{i-l}}} \right]\right\} \max\left\{0, V_l \left[y_{t_{i-l}} - \mu_{y|x_{t_{i-l}}} \right]\right\}^{'}$$

$$+ \sum_{l=1}^{L_\omega} W_l \, x_{(1),t_{i-l}} x_{(1),t_{i-l}}^{'} W_l^{'}.$$

where $x_{(1),t_{i-l}}$ indicates that only the first column of $x_{t_{i-l}}$ enters the computation, the $\max(0,x)$ function is applied element wise.[30] The matrices P_l, Q_l, V_l , and W_l can be scalar, diagonal, or full M by M matrices and the number of lags to compute R_x is $\max\left(L_r + L_u, L_v + L_u, L_\omega\right)$. Summarizing the parameter of interest are in equation (25).

$$\theta_{\mathcal{T}} = \mathrm{vec}\left[a_{(\cdot)(\mathbf{0})} \,|\, A \,|\, \mathfrak{b}_0 \,|\, B \,|\, R_0 \,|\, Q_1 \ldots Q_{L_g} \,|\, P_1 \ldots P_{L_r} \,|\, V_1 \ldots V_{L_v} \,|\, W_1 \ldots W_{L_\omega} \right] \tag{25}$$

Table 1 summarizes some possible settings and their relation with known models.[31]

Table 1: Different specifications and their implied models

Model	L_u	L_g	L_r	K_z	K_x	L_p
Gaussian iid	$= 0$	$= 0$	$= 0$	$= 0$	$= 0$	≥ 0
Gaussian VAR	> 0	$= 0$	$= 0$	$= 0$	$= 0$	≥ 0
Semiparametric VAR	> 0	$= 0$	$= 0$	> 0	$= 0$	≥ 0
Gaussian ARCH	≥ 0	$= 0$	> 0	$= 0$	$= 0$	≥ 0
Semiparametric ARCH	≥ 0	$= 0$	> 0	> 0	$= 0$	≥ 0
Gaussian GARCH	≥ 0	> 0	> 0	$= 0$	$= 0$	≥ 0
Semiparametric GARCH	≥ 0	> 0	> 0	> 0	$= 0$	≥ 0
Nonlinear nonparametric	≥ 0	> 0	> 0	> 0	> 0	≥ 0

L_u is the number of lags to compute $\mu_{y|x}$. L_g is the number of terms in the autoregressive part that involves Q_l and L_r is the number of terms in the moving average part that involves P_l in equation (24). K_z is the degree of the polynomial in z, K_x is the degree of the polynomial to construct the coefficients a_x and L_p is number of lags to compute the coefficients a_x as polynomials (if $K_x = 0$ then any $L_p > 0$ has no effect).

The parameters in the terms of the polynomials involving interactions might grow to fast for the available sample and in that case it is useful to control its quantity in the estimation procedure. Thus, the

---

[30]$\sum_{y|x_{t_i}}$ must be differentiable with respect to the parameters in $\mu_{y|x_{t_{i-l}}}$, so the $\max\{0,x\}$ function actually applied is a twice continuously differentiable cubic ne approximation. The approximation agrees with the $\max\{0,x\}$ function except over the interval $(0, 0.1)$ where it lies just above the $\max\{0,x\}$ function.

[31]Table from Gallant and Tauchen (2013b)

maximum degree of interactions included in the polynomial on z and x are I_z and I_x respectively. The coefficients a_x are allowed to be polynomials of degree K_x up to those coefficients included in the terms of degree $maxK_z$ of the polynomial on z. Similarly, those coefficients included in the cross terms of degree higher than $maxI_z$ do not depend on x.

Table 2 displays the best fit for the SNP density in bold along with some other possible specifications roughly sorted by groups of models. The sample is a collection of 3185 daily observations of sovereign CDS and Bond prices at 5 years of maturity for Brazil (from 10-12-2001 to 5-22-2014). The details about the collected data and the construction of the sample are in Appendix C.

Table 2: Model selection of the SNP auxiliary model

																Free Par.	Obj. Func.	BIC
L_u	L_g	$type\ Q$	L_r	$type\ P$	L_v	$type\ V$	L_w	$type\ W$	L_p	K_z	I_z	$\max K_z$	$\max I_z$	K_x	I_x			
1	0	s	0	s	0	s	0	s	1	0	0	0	0	0	0	9	-2.51322	-2.50182
2	0	s	0	s	0	s	0	s	1	0	0	0	0	0	0	13[i]	-2.54495	-2.52849
3	0	s	0	s	0	s	0	s	1	0	0	0	0	0	0	17	-2.54686	-2.52533
2	0	s	0	s	0	s	0	s	1	4	0	4	0	0	0	21	-3.00094	-2.97435
2	0	s	0	s	0	s	0	s	1	5	0	5	0	0	0	23	-3.03083	-3.00171
2	0	s	0	s	0	s	0	s	1	6	0	6	0	0	0	25	-3.0587	-3.02704
2	0	s	0	s	0	s	0	s	1	7	0	7	0	0	0	27	-3.10009	-3.0659
2	0	s	0	s	0	s	0	s	1	8	0	8	0	0	0	29	-3.11321	-3.07649
2	0	s	0	s	0	s	0	s	1	9	0	9	0	0	0	31	-3.12616	-3.0869
2	0	s	0	s	0	s	0	s	1	10	0	10	0	0	0	33	-3.14847	-3.10669
2	0	s	0	s	0	s	0	s	1	11	0	11	0	0	0	35	-3.1514	-3.10708
2	0	s	0	s	0	s	0	s	1	12	0	12	0	0	0	37	-3.16854	-3.12169
2	0	s	0	s	0	s	0	s	1	13	0	13	0	0	0	39	-3.17487	-3.12549
2	0	s	0	s	0	s	0	s	1	14	0	14	0	0	0	41	-3.18418	-3.13226
2	0	s	0	s	0	s	0	s	1	15	0	15	0	0	0	43	-3.18844	-3.13399
2	0	s	0	s	0	s	0	s	1	16	0	16	0	0	0	45	-3.19606	-3.13907
2	0	s	0	s	0	s	0	s	1	17	0	17	0	0	0	47	-3.19771	-3.13819
2	0	s	0	s	0	s	0	s	1	12	2	12	0	0	0	38	-3.18558	-3.13746
2	0	s	0	s	0	s	0	s	1	10	3	10	0	0	0	36	-3.17097	-3.12539
2	0	s	0	s	0	s	0	s	1	16	2	16	0	0	0	46	-3.24752	-3.18928
2	0	s	0	s	0	s	0	s	1	16	3	16	0	0	0	48	-3.25784	-3.19706
3	0	s	0	s	0	s	0	s	1	16	2	16	0	0	0	50	-3.26269	-3.19938
2	0	s	0	s	0	s	0	s	1	16	4	16	0	0	0	51	-3.27875	-3.21417
2	0	s	0	s	0	s	0	s	1	16	2	16	2	0	0	46	-3.25177	-3.19352
2	0	s	0	s	0	s	0	s	1	16	4	16	3	0	0	51[ii]	-3.29187	-3.22728
2	0	s	1	s	0	s	0	s	1	0	0	0	0	0	0	14	-2.79193	-2.7742
2	1	s	1	s	0	s	0	s	1	0	0	0	0	0	0	15	-3.20641	-3.18742
3	1	s	1	s	0	s	0	s	1	0	0	0	0	0	0	19	-3.20746	-3.1834
2	2	f	1	s	0	s	0	s	1	0	0	0	0	0	0	22	-3.22276	-3.1949
2	1	s	2	f	0	s	0	s	1	0	0	0	0	0	0	22	-3.30235	-3.27449
2	2	f	2	f	0	s	0	s	1	0	0	0	0	0	0	29[iii]	-3.33227	-3.29554
2	2	f	2	f	0	s	1	f	1	0	0	0	0	0	0	33	-3.39301	-3.35123
2	2	f	2	f	0	s	0	s	1	4	0	4	0	0	0	37	-3.52458	-3.47773
2	2	f	2	f	0	s	0	s	1	4	2	4	0	0	0	38	-3.5477	-3.49959
2	2	f	2	f	1	f	0	s	1	4	2	4	0	0	0	42	-3.55135	-3.49816
3	2	f	2	f	0	s	0	s	1	4	2	4	0	0	0	42	-3.55277	-3.49958

												Model				Free Par.	Obj. Func.	BIC
L_u	L_g	type Q	L_r	type P	L_v	type V	L_w	type W	L_p	K_z	I_z	max K_z	max I_z	K_x	I_x			
2	2	f	2	f	0	s	0	s	1	5	2	5	0	0	0	40	-3.54827	-3.49762
2	1	s	2	f	0	s	0	s	1	5	2	5	0	0	0	33	-3.53025	-3.48847
2	2	f	2	f	0	s	1	f	1	4	2	4	0	0	0	42	-3.61949	-3.5663
2	2	f	3	f	0	s	1	f	1	4	2	4	0	0	0	46	-3.62064	-3.56239
2	2	f	2	f	0	s	1	f	1	4	2	4	2	0	0	42^{iv}	-3.62099	-3.56781
2	2	f	2	f	1	s	1	f	1	4	2	4	0	0	0	43	-3.62198	-3.56753
2	2	f	2	f	0	s	1	f	1	4	3	4	3	0	0	44	-3.6227	-3.56698
2	0	s	0	s	0	s	0	s	2	16	0	16	0	1	2	177	-3.27952	-3.05539
2	2	f	2	f	0	s	1	f	1	4	2	4	0	1	2	60	-3.65222	-3.57624
2	2	f	2	f	0	s	1	f	2	4	2	4	0	1	2	78	-3.66095	-3.56218
2	2	f	2	f	0	s	1	f	1	4	3	4	0	1	2	62	-3.65452	-3.57601
2	3	f	2	f	0	s	1	f	1	4	2	4	0	1	2	64	-3.65263	-3.57159
2	2	f	2	f	0	s	1	f	1	4	2	4	0	2	2	87	-3.66779	-3.55762
2	2	f	2	f	0	s	1	f	1	5	2	5	0	1	2	66	-3.6614	-3.57783
2	2	f	2	f	1	d	1	f	1	4	2	4	0	1	2	62	-3.65277	-3.57426
2	2	f	2	f	0	s	2	f	1	4	2	4	0	1	2	64	-3.65418	-3.57314
2	2	f	2	f	0	s	1	f	1	5	3	5	0	1	2	68	-3.6619	-3.57579
2	2	f	2	f	0	s	1	f	1	5	5	5	4	1	2	87	-3.69594	-3.58578
2	2	f	2	f	0	s	1	f	1	6	2	6	6	1	2	74	-3.68205	-3.58834
2	2	f	2	f	0	s	1	f	1	7	2	7	5	1	2	80	-3.68536	-3.58405
2	2	f	3	f	0	s	1	f	1	6	2	6	5	1	2	78	-3.68291	-3.58414
3	2	f	2	f	0	s	1	f	1	6	2	6	6	1	2	78	-3.68396	-3.58519
2	3	f	2	f	0	s	1	f	1	6	2	6	6	1	2	78	-3.68305	-3.58428
2	2	f	2	f	1	s	1	f	1	6	2	6	6	1	2	75	-3.68246	-3.58749
2	2	f	2	f	0	s	2	f	1	6	2	6	6	1	2	78	-3.68275	-3.58398
2	2	f	2	f	0	s	1	f	2	6	2	6	6	1	2	102	-3.6916	-3.56244
2	2	f	2	f	0	s	1	f	1	6	3	6	6	1	2	80	-3.68752	-3.58621
2	2	f	2	f	0	s	1	f	1	6	2	6	6	2	2	116	-3.70059	-3.5537
2	2	f	2	f	0	s	1	f	1	6	2	6	2	1	1	**74^{v}**	**-3.68205**	**-3.58834**
2	2	f	2	f	0	s	1	f	1	6	2	6	2	2	1	102	-3.69918	-3.57002
2	2	f	2	f	0	s	1	f	1	6	3	6	3	1	1	80	-3.68752	-3.58621
2	2	f	1	f	0	s	1	f	1	7	2	7	2	1	1	76	-3.67489	-3.57866

Superscripts i, ii, iii, iv, v indicate the best fit for a Gaussian VAR, seminonparametric VAR, Gaussian GARCH, seminonparametric GARCH and nonlinear SNP respectively. The BIC in this table is Obj. Func$+k \log(n)/2n$.

The best fit for the auxiliary model corresponds to a nonlinear SNP specification with 74 parameters. The specification includes 2 lags to model the conditional mean and three of the four types of matrices that incorporate leverage and level effects in the variance function have been useful to improve the BIC. The degrees of the polynomials in z and x are 6 and 1 respectively.

IV.2 Reduced form parameters for different models

The parameters of the reduced form model define the stochastic differential equation for the intensity of default in equation (26).

$$d \ln \lambda_t = b_\rho \left(\ln \lambda_t \right) dt + \sigma_\rho \left(\ln \lambda_t \right) dW_t \tag{26}$$

Two models that are interesting to estimate as benchmarks are the well known Ornstein–Uhlenbeck (OU) and Cox-Ingersoll-Ross[32] (CIR) processes in equations (27) and (28). These processes are broadly used in applications in part because they allow the derivation of closed form solutions for the survival probability, which greatly simplifies the computation of risky asset prices. Note that the estimation strategy proposed in this paper is meant to work for any ρ that defines a stationary process and does not take advantage of these particular features of the OU and CIR process to estimate their parameters.

The OU model for the intensity of default is

$$d \ln \lambda_t = \beta_0 \left(\beta_1 - \ln \lambda_t \right) dt + \gamma_0 \, dW_t \tag{27}$$

where for equation (27), $\beta_0, \gamma_0 > 0$.

The CIR for the intensity process follows

$$d\lambda_t = \beta_0 \left(\beta_1 - \lambda_t \right) dt + \gamma_1 \sqrt{\lambda_t} \, dW_t \tag{28}$$

where $\beta_0, \gamma_1 > 0$ and $2\beta_0\beta_1 > \gamma_1^2$ to ensure a strictly positive process in equation (28).

Table 3 displays the estimates for the OU and CIR processes. For the OU process $\rho_0 = \beta_0\beta_1$ and $\rho_1 = \beta_0$, and for the CIR process $\rho_0 = \beta_0(\beta_1 - \frac{\gamma_1^2}{2\beta_0})$ and $\rho_1 = \beta_0$. The estimates for equation (28) correspond to the log transform[33] of the process, i.e. $d \ln \lambda_t = (\rho_0 \lambda_t^{-1} - \rho_1) \, dt + \gamma_1 \lambda_t^{-\frac{1}{2}} \, dW_t$.

Table 3: Estimates for the OU and CIR processes

ρ	OU			CIR		
	Mean	**Mode**	$\mathbf{SE}_H$	**Mean**	**Mode**	$\mathbf{SE}_H$
ρ_0	-0.69263	-0.7263[***]	0.04317	0.00053	0.0009[***]	0.00028
ρ_1	0.11210	0.1069[***]	0.03546	0.43361	0.4430[***]	0.02401
γ_0	0.23633	0.2171[***]	0.06032			
γ_1				0.04438	0.0443[***]	0.00385

Degrees of freedom: $d(\theta) - 1 - d(\rho) = 71$. Sample size: 3185. Period: 10-12-2001 to 5-22-2014 (daily). The objective function, $s(\rho) = m'_\mathcal{T}(\rho, \hat{\theta}_\mathcal{T})\hat{\mathcal{J}}_\mathcal{T}^{-1}m_\mathcal{T}(\rho, \hat{\theta}_\mathcal{T})$, at the mode is 0.49 for the OU and 0.51 for the CIR. [***] if p-value < 0.01, [**] if p-value $\in [0.01, 0.05]$ and [*] if p-value $\in (0.05, 0.1]$.

Considering the values of the objective functions displayed in table 3, model OU performs better than the CIR. Thus, as an alternative model related to the OU process consider for instance Model "A" given in equations (29) to (33). This model nests the OU process for all values in $[\ln \underline{\lambda}_t, \ln \overline{\lambda}_t]$. The diffusion coefficient in model A always takes positive values and behaves as $\psi_\gamma(\cdot)$ indicates in equation (31) when its value is above a lower bound $\underline{\sigma}$ otherwise $\sigma_\rho(\cdot) = \underline{\sigma}$. There are no kinks in $\sigma_\rho(\cdot)$ because the pseudo-indicator function in equation (32) controls the smoothness of the transition. To avoid stationarity issues the drift behaves as described by $\psi_\beta(\cdot)$ in equation (30) except for extreme values of the intensity where

---

[32] Also known as a square-root process.

[33] See Appendix B for a note on this standard transformation.

the drift becomes a constant as indicated by $b_\rho\left(\cdot\right).$[34]

$$b_\rho\left(\ln\lambda_t\right) = \begin{cases} \psi_\beta\left(\ln\underline{\lambda}_t\right) & \text{if } \lambda_t < \underline{\lambda}_t \\ \psi_\beta\left(\ln\lambda_t\right) & \text{if } \lambda_t \in [\underline{\lambda}_t, \overline{\lambda}_t] \\ \psi_\beta\left(\ln\overline{\lambda}_t\right) & \text{if } \lambda_t > \overline{\lambda}_t \end{cases} \tag{29}$$

$$\psi_\beta\left(\ln\lambda_t\right) = \beta_0\left(\beta_1 - \ln\lambda_t\right) + \beta_2\left(\frac{1}{2} + \lambda_t\right)^{-3} + \beta_3\left(\ln\lambda_t\right)^2 \tag{30}$$

$$\sigma_\rho\left(\ln\lambda_t\right) = \psi_\gamma\left(\ln\lambda_t\right) I_{\underline{\sigma}}\left(\ln\lambda_t\right) + \frac{\sigma}{2}\left(1 - I_{\underline{\sigma}}\left(\ln\lambda_t\right)\right)$$

$$\psi_\gamma\left(\ln\lambda_t\right) = \gamma_0 + \gamma_1\left(\frac{1}{2} + \lambda_t\right)^{\frac{3}{2}} \tag{31}$$

$$I_{\underline{\sigma}}\left(\ln\lambda_t\right) = \frac{1}{1 + e^{[\underline{\sigma} - \psi_\gamma(\ln\lambda_t)]\alpha}} \tag{32}$$

$$\ln\overline{\lambda}_t = 6,\ \ln\underline{\lambda}_t = -26,\ \underline{\sigma} = 0.1^3,\ \alpha = 1\times10^3. \tag{33}$$

Unlike the drift and the diffusion of an OU process, the specifications of Model A can take a variety of nonlinear shapes which might help to fit the observations better. Table 4 displays the estimates for model A.[35]

Table 4: Estimates for model A

ρ	Mean	Mode	SE_H
ρ_0	-0.38772	-0.3977***	0.01211
ρ_1	0.19068	0.2004***	0.00718
ρ_2	-0.14303	-0.1406***	0.00221
ρ_3	0.01319	0.0136***	0.00073
γ_0	0.52965	0.5137***	0.01392
γ_1	-0.60488	-0.5711***	0.02518

Degrees of freedom: $d(\theta) - 1 - d(\rho) = 68$. The objective function at the mode is 0.40. *** if p-value < 0.01, ** if p-value $\in [0.01, 0.05]$ and * if p-value $\in (0.05, 0.1]$.

The fit of models A and OU cannot longer be compared directly through the values of the corresponding objective functions. The BICs[36] for models A and OU are approximately 2580 and 3125 respectively, thus after taking into account a penalty due to the additional parameters Model A still performs better than model OU.

Figure 1 presents a comparison between the drift and the diffusion coefficients evaluated at the estimates of both models, the OU model as the "benchmark" and Model A as the "fit".

---

[34]The kinks in $b_\rho\left(\cdot\right)$ can be avoided but they are irrelevant in the estimation because $\ln\overline{\lambda}_t$ and $\ln\underline{\lambda}_t$ are the boundaries of the support.

[35]Any given model was started from different initial conditions and its coefficients reached similar values of the same sign. For instance, for model A the average of the absolute values of the differences between coefficients for different initial conditions was 0.06. This average is smaller for model OU and CIR and can be reduced further increasing the amount of time that the models are allowed to run.

[36]For a number of parameters k, $BIC = -2\log(e^{-ns(\rho)}) + k\log(T)$.

Figure 1: Functional Forms: OU (benchmark) and Model A (fit)

(a) $b_\rho(\ln \lambda_t)$

(b) $\sigma_\rho(\ln \lambda_t)$
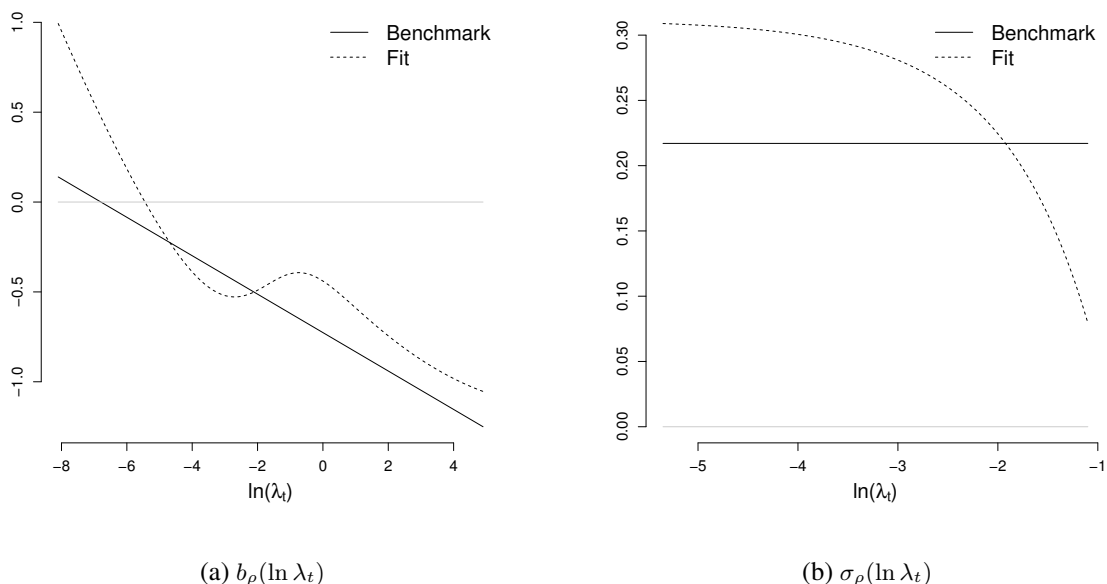
Disregarding the stochastic component of the model, the drift function of Model A has a higher equilibrium point than the benchmark, which makes the interpretation of the results more intuitive since the equilibrium point would be reached at an intensity level not so close to 0. Other feature of model A is the different speeds at which the drift adjusts a disequilibrium. If the intensity parameter is off equilibrium by a small amount, the drift adjusts a positive or negative disequilibrium at similar rates. If the intensity is off equilibrium by a larger positive amount, the speed of adjustment flattens and becomes even lower than the estimates for the OU process. Finally, if the intensity of default becomes too high, the drift for model A imposes higher rates of adjustment towards a lower intensity. These features can be an attempt to improve the fit to higher CDS prices which might take longer to be reduced than what would be suggested by a linear model. In the case of the diffusion function, its values remain above the OU process' estimates for all intensity levels but the extremely high ones. It is intuitive to think that for unusually high intensity levels the process is mainly driven by the drift, which pulls the intensity towards lower levels.

Summarizing, it has been possible to find an alternative model, Model A, which performs better than the usual benchmarks. Moreover, other alternative models for the pricing equations can be proposed and the same estimation approach would be useful because it does not depend on particular reduced form models or specifications of the intensity process. As a result, this estimation approach does not restricts the researcher's models to those known to have a closed form solution or an approximation of the unknown likelihood.

V   Conclusions

Diffusion processes have been broadly used to model the prices of financial assets, but most of them are particularly hard to estimate because their law of motion is unknown in closed form. On top of this, diffusions are useful to model the trajectories of unobservable parameters, such as the intensity of default that is involved in the pricing of financial assets with credit risks. Thus, complementary challenges emerge when the objective of the researcher is estimating alternative diffusion processes for the intensity of default to improve the modeling of risky asset prices.

This paper proposes a new approach combining a semi-nonparametric estimator in the framework of the Efficient Method of Moments, Gallant and Tauchen (1996) and a reduced-form model for pricing sovereign bonds and Credit Default Swaps. The intensity of default is an unobservable process, there are not enough observations of credits events for a country to estimate its default probabilities through a direct approach, and in addition the historical measure is not the right one for asset pricing. Hence, the reduced form model provides a map between the unobservable intensity and the observable risky asset prices that, under a suitable estimator, allows to estimate the parameters for the intensity process. For general specifications, the likelihood function of this model is unknown in closed form in part because most of the diffusions have unknown law of motions. The estimation strategy solves this issue through an EMM estimator and a semi-nonparametric auxiliary function that provides the building blocks to construct its moment conditions. One desirable feature of the EMM is that under some assumptions (beyond the required for consistency), the estimator is as efficient as maximum likelihood.

The reduced form model suggested in this paper is a departure from the model in Pan and Singleton (2008). For instance, this paper proposes a reduced form model that requires data on sovereign CDS prices for some maturity and the underlying prices of the bonds, while the model in Pan and Singleton (2008) requires data on sovereign CDS for several maturities,[37] this difference is particularly useful when the data on sovereign CDS for several maturities is scarce. In addition, the model in this paper allows for arbitrary distributions of the pricing errors, and this is an advantage with respect to the alternative models with some maturity of the CDS pricing exactly and other maturities pricing with errors that follow a know type of distribution.

The application for sovereign CDS and Bond prices of Brazil estimates the reduced form model under 3 specifications of the diffusion process for the intensity parameter, the well known Ornstein–Uhlenbeck (OU) and Cox-Ingersoll-Ross (CIR) processes as benchmark models and an alternative process, "Model A", that to the best of my knowledge does not provide closed form solutions for the survival probabilities.[38] Considering the values of the Bayesian information criterion, this paper finds that Model A performs better than the benchmark models.

This paper contributes to the estimation of more flexible models for risky asset prices and estimates an original model for the intensity of default that performs better than the known OU and CIR processes. The estimation strategy works with different reduced form models with unknown distributions of the pricing errors and likelihood functions that are difficult or not possible to derive.

---

[37]Both models also use data to construct risk free yield curves.

[38]The estimation strategy is meant to work for any specification that defines a stationary process and does not take advantage of the particular features of the OU and CIR process to estimate their parameters.

Appendix A Stationary Diffusion

Let $Z_t = \ln \lambda_t$ be a diffusion process described by equation (2) with domain $\mathcal{D}$. Given z_0, the solution of equation (2) exists and it is unique if b and σ are functions satisfying certain regularity conditions. Assumptions A1 to A3 provide sufficient conditions to ensure that these regularity conditions are satisfied.

Assumption A1. *For all $z, y \in \mathcal{D}$, there exists a $K \in (0, \infty)$ such that b and σ are measurable functions that satisfy :*

$$\|\sigma(z) - \sigma(y)\| + \|b(z) - b(y)\| \leq K \|z - y\|$$

where $\|\cdot\|$ denote the the Euclidian norm.

Given an Ito diffusion Z_t with initial condition z, an Ito integral is a martingale with respect to the probability measure P_z induced by the law of Z_t when $Z_0 = z$. If for some given initial condition z and a pair $\left\{ b(\cdot), [\sigma(\cdot)]^2 \right\}$ there is a unique probability measure that solves the associated martingale problem,[39] then the martingale problem is said to be well posed. If the martingale problem is well posed for all initial conditions, then it can be proved that there is a weak solution to equation (2) and such solution is a Markov process (Rogers and Williams 2000, Stroock and Varadhan 2005). assumptions A1 and A2 are sufficient for the existence of a strong solution of equation (2) (for a general initial condition z_0, Karatzas and Shreve 1991 p. 289) and ensure its weak uniqueness.[40]

Assumption A2. *For all $z \in \mathcal{D}$, there exists a $K \in (0, \infty)$ such that*

$$\|\sigma(z)\|^2 + \|b(z)\|^2 \leq K^2 \left(1 + \|z\|^2 \right).$$

Assumption A3 becomes useful to construct a diffusion with the mentioned properties that is associated to (an operator defined through) a pair $\left\{ b(\cdot), [\sigma(\cdot)]^2 \right\}$.

Assumption A3. *For all $z \in \mathcal{D}$, $[\sigma(z)]^2$ is positive definite.*

Assumptions A1 to A3 give sufficient conditions to ensure that the SDE in equation (2) has a solution for all λ_0 and that the martingale problem is well-posed starting at λ_0, so there is a family of unique probability measures $\{P_z, : z \in \mathcal{D}\}$, one for each initial condition, induced by the solutions to equation (2). Then under each P_z, $\{Z_t : t \geq 0\}$ is a time-homogeneous strong Markov process (see Rogers and Williams 2000 p. 162-163). These assumptions can also be replaced by alternative conditions that ensure these properties (see for instance Øksendal 2000, Rogers and Williams 2000, Stroock and Varadhan 2005).

If a diffusion has an invariant distribution, and its initial condition is drawn from its invariant distribution then it is stationary. To ensure that a diffusion has an invariant distribution it is sufficient to consider positive recurrent diffusions and the following concepts are useful to define them. The scale function of a diffusion Z is given in equation (34).

$$s(z) = \int_{y_0}^{z} \exp \left(- \int_{x_0}^{y} \frac{2b(x)}{\sigma^2(x)} dx \right) dy, \tag{34}$$

---

[39] A definition of the martingale problem can be found in Definition 8.3.2 of Øksendal (2000).

[40] A strong solution is also weak, but the converse is not true. A weak solution is unique if any other solution is identical in law, i.e. have the same finite-dimensional distribution.

where y_0, x_0 are arbitrary points in $\mathcal{D}$. The speed density is

$$m(z) = \left[\sigma^2(z)\, s'(z)\right]^{-1}, \tag{35}$$

where s' is the derivative of s. Let τ_z be the stopping time at some z in $\mathcal{D}$, thus

$$\tau_z = \inf\{t \geq 0 | Z_t = z\}.$$

Then, the diffusion Z is recurrent if for all $z, y \in \mathcal{D}$,

$$\Pr\{\tau_z < \infty | Z_0 = y\} = 1.$$

If b and σ satisfy the regularity conditions and the scale function is unbounded, i.e.

$$\lim_{z \to \pm\infty} s(z) = \pm\infty, \tag{36}$$

then the diffusion Z is recurrent. When for all $z, y \in \mathcal{D}$,

$$E\left[\tau_z < \infty | Z_0 = y\right] = \infty$$

the diffusion Z is null recurrent, and when

$$E\left[\tau_z < \infty | Z_0 = y\right] < \infty$$

the diffusion Z is positive recurrent.[41] A diffusion Z is null recurrent if

$$\int_{\mathcal{D}} m(z)\, dz = \infty,$$

and it is positive recurrent if

$$\int_{\mathcal{D}} m(z)\, dz < \infty. \tag{37}$$

Assumption A4 summarizes the additional sufficient conditions to ensure that the diffusion is positive recurrent.

Assumption A4. *The scale function s of the diffusion Z is unbounded and its speed measure is finite, i.e. equations (36) and (37) hold.*

Note that for one dimensional diffusions, the invariant density of Z, μ', if exists, must be proportional to the speed density m, i.e. $\mu'(z) \propto m(z)$.

Appendix B   Log transform of an Ito diffusion

The programs for the simulations are written to handle diffusions for $\ln \lambda_t$. Thus, for instance in case of the CIR process the user can define the analogous diffusion in logs as follows. By Ito calculus, for any twice differentiable scalar function $f(t, x_t)$ of real variables t and x,

$$df(t, x_t) = \frac{\partial f(\cdot)}{\partial t} dt + \frac{\partial f(\cdot)}{\partial x_t} dx_t + \frac{1}{2} \frac{\partial^2 f(\cdot)}{\partial x_t{}^2} dx_t{}^2$$

---

[41]If a diffusion is not recurrent, then it is transient. Transient and null recurrent diffusions are not stationary.

Thus for $\ln \lambda_t$,

$$d \ln \lambda_t = \frac{1}{\lambda_t} d\lambda_t + \frac{1}{2} \left(\frac{-1}{\lambda_t^2} \right) d\lambda_t^2$$

and by equation (28), it follows

$$d \ln \lambda_t = \frac{1}{\lambda_t} d\lambda_t - \frac{1}{2\lambda_t^2} \left(\beta_0 (\beta_1 - \lambda_t) dt + \gamma_1 \sqrt{\lambda_t} \, dW_t \right)^2$$

then by the Ito calculus $dW_t dt = 0$, $(dt)^2 = 0$ and $(dW_t)^2 = dt$ so

$$d \ln \lambda_t = \frac{1}{\lambda_t} d\lambda_t - \frac{\gamma_1^2}{2} \lambda_t^{-1} dt.$$

Replacing $d\lambda_t$ with equation (28),

$$d \ln \lambda_t = \beta_0 \left(\lambda_t^{-1} \left(\beta_1 - \frac{\gamma_1^2}{2\beta_0} \right) - 1 \right) dt + \gamma_1 \lambda_t^{-\frac{1}{2}} \, dW_t. \tag{38}$$

Note that in general for $d\lambda_t = \mu(\lambda_t) dt + \sigma(\lambda_t) dW_t$ it is true that $d \ln \lambda_t = \left(\frac{\mu(\lambda_t)}{\lambda_t} - \frac{\sigma^2(\lambda_t)}{2\lambda_t^2} \right) dt + \frac{\sigma(\lambda_t)}{\lambda_t} dW_t$. Whenever is more numerically stable, the program can automatically transform a diffusion specified in logs,[42] simulate from the corresponding diffusion in levels and then transform the paths back to their logs.

Appendix C   Data description

The zero-coupon fixed-income yield curve is constructed using the London Interbank Offered Rate (Libor) for U.S. dollars, i.e. the average interest rate at which leading banks borrow funds from other banks in the London market, and the "swaps rates" at different maturities, i.e. the mid-market par rate paid by a fixed-rate payer on an interest rate swap agreement with certain maturity in return for receiving three month LIBOR. The data was collected from the website of the Federal Reserve Bank of St. Louis, the source of the Libor rate is ICE Benchmark Administration Limited (IBA) and the source for the swaps rates is the Board of Governors of the Federal Reserve System. The maturity of the collected rates are as follows: overnight, 1 week, 1, 2, 3, 6 and 12 months for the Libor rates and 1, 3, 4, 5, 7 and 10 years for the swap rates. The missings in the overnight Libor rates and in the swaps rates at 1 year are imputed according to the growths observed in the Libor rates at 1 week and 12 moths respectively, while the remaining missings in the collected data are interpolated.[43] To obtain rates at regular intervals, a cubic spline interpolation and a linear interpolation are adjusted across the rates for different maturities available per date and their results are averaged. Finally, the standard procedure known as "bootstrapping" is applied to calculate the zero-coupon yield curve, and starting from the swap rates of more than one year of maturity the calculation is recursive. Table C1 summarizes the means and standard deviations (sd) by maturities of averages across the bootstrapped zero-coupon yield curves.

---

[42]So far the program accepts a drift and diffusion coefficients for a process specified in logs only and there is a switch to transform them to levels.

[43]146 observations per series in a database of 3494 were imputed by direct interpolation.

Table C1: Statistics by maturities of averages across zero-coupon yield curves

| Mat. | Curves' averages | | | | | | | |
| | 2001 to 2004 | | 2005 to 2008 | | 2009 to 2012 | | 2012 to 2014 | |
	mean	sd	mean	sd	mean	sd	mean	sd
1	2.185	0.085	4.272	0.089	0.569	0.159	0.312	0.074
2	2.647	0.169	4.310	0.009	0.677	0.061	0.353	0.041
3	3.239	0.168	4.378	0.027	1.025	0.125	0.590	0.088
4	3.721	0.110	4.474	0.027	1.417	0.098	0.917	0.099
5	4.058	0.087	4.562	0.024	1.740	0.092	1.258	0.099
6	4.320	0.065	4.640	0.020	2.033	0.077	1.581	0.086

Prepared based on Libor and swaps rates published by IBA and the Board of Governors of the Federal Reserve System respectively.

The data for the CDS prices corresponds to the last price traded and was collected from different sources that publish the quotes through Bloomberg. The sources are CBIN, CBIL, CBGL, CBBT, CBIT, HCDS, HELC, CBGT, RCDS, CBGL, CBGN, DZCD, CMAN, CMAI, CMAQ and CMAL. The currency of the CDS is U.S. dollars, the pricing unit is basis points, the coupon frequency is quarterly, their recovery rate is 0.25 and their maturity is 5 years. To avoid possible outliers, if some of the individual quotes exceeds the mean of the quotes in the corresponding day by more than 2.58 times the standard deviation (calculated over a moving window of 25 working days), the quote is dropped from the sample.

The bond prices are the daily last price in the "USD Brazil Sovereign curve" of Bloomberg at 5 years. The constituents of the index are tracked to obtain the coupon rates and the exact maturity of the corresponding bonds. The missing observations in the index are imputed according to the growth of a series of bond prices calculated using yield to maturities, coupon rates and maturities imputed as follows. When available, the yield to maturity, coupon rates and the exact maturity agree with the ones for the underlying bonds in the "USD Brazil Sovereign curve" of Bloomberg at 5 years. The missing values for the yield to maturity are completed interpolating the series as long as the bond id remains the same. When possible, the remaining missing values of the yields are imputed according to the average of the growth rates for the predicted yields in the following 3 linear regressions. The explanatory variables of the first regression are functions of the Anbima Brazil Government Bond Fixed Rate 5 Years yield for bonds in local currency, the yield to maturity of the underlying bonds in the Bloomberg's "USD Brazil Sovereign curve" for 2 years of maturity, the coupons rates for the underlying bonds in USD at 5 and 2 years of maturity and the exchange rate Real–US Dollar. The explanatory variables of the second regression are functions of the yield to maturity of the underlying bonds in the Bloomberg's "USD Brazil Sovereign curve" for 8 years of maturity and the coupons rates for the underlying bonds in USD at 5 and 8 years of maturity. The explanatory variables of the third regression are functions of the yield to maturity of the Brazilian sovereign bond in USD issued on 14 of November of 2006 and maturing on 17 of January of 2017, the differences of coupon rates between the just mentioned bonds and the underlying bonds in the USD Brazil Sovereign curve at 5 years of maturity. For those missing yields that could be imputed, the missing values for the maturities are interpolated and the missing values for the coupon rates are assumed to be equal to the last available coupon rate. A series of bond prices is calculated (for a face value of 100 US dollars) using the compiled and imputed data for the yields, maturities and coupons. Finally, the growth rates of this calculated series of bond prices is used to impute missing values in the series of USD

Brazil Sovereign bond prices at 5 years collected from Bloomberg.

Table C2: Statistics by year for the USD Brazil Sovereign bond and CDS prices at 5 years

| Year | CDS | | | | Bonds | | | | Mat. | | Coupon | |
	mean	min	max	sd	mean	min	max	sd	mean	sd	mean	sd
2001	948.75	807.50	1117.50	112.86	92.94	86.73	96.86	3.09	4.15	0.06	0.10	0.000
2002	1909.58	702.50	3951.50	1059.78	79.96	51.37	100.99	16.06	4.58	0.53	0.11	0.005
2003	995.60	402.10	2152.50	429.60	98.07	77.64	109.64	8.42	4.51	0.18	0.10	0.009
2004	543.76	302.33	902.69	148.69	106.34	95.58	112.19	4.16	3.76	0.29	0.09	0.000
2005	327.96	221.01	452.14	53.84	118.79	109.05	123.74	3.82	4.46	0.60	0.12	0.010
2006	141.82	98.75	225.07	30.14	119.01	113.53	124.51	3.19	4.60	0.42	0.11	0.010
2007	89.64	61.59	155.63	17.49	120.23	117.15	123.64	1.95	4.40	0.18	0.11	0.005
2008	187.69	86.13	593.02	105.26	121.31	104.68	126.08	4.42	4.76	0.40	0.10	0.003
2009	211.76	110.84	427.86	97.20	124.54	116.50	129.94	3.45	4.93	0.30	0.10	0.006
2010	121.97	89.73	153.56	13.32	119.84	116.00	124.65	2.37	4.68	0.29	0.08	0.000
2011	132.65	100.21	217.81	28.25	120.26	113.50	127.32	3.84	4.99	0.38	0.07	0.009
2012	130.02	99.67	175.43	19.14	117.90	116.48	119.76	0.77	5.01	0.04	0.06	0.000
2013	156.97	101.41	215.10	34.20	115.50	111.97	119.20	2.47	4.88	0.04	0.06	0.000
2014	173.95	144.03	210.07	18.76	113.41	112.14	114.58	0.70	4.77	0.05	0.06	0.001

Prepared based on Bloomberg's data on the USD Brazil Sovereign curve and CDS prices.

Appendix D   Computing expectations by its Feynman-Kac representation applying the Crank-Nicolson method

Let X be an Ito process starting at time t_0 at the point x. Let $t_i \in [t_0, t]$ such that

$$X_t = x + \int_{t_0}^{t} b(X_s) \, ds + \int_{t_0}^{t} \sigma(X_s) \, dW_s. \tag{39}$$

Define the function $v : \mathbb{R}^K \times [t_0, t] \to \mathbb{R}$ by

$$v(x, t) = E\left[e^{-\int_{t_0}^{t} \varphi(X_s) ds} g(X_t) \,|\, X_0 = x \right], \tag{40}$$

where g and φ are a twice continuously differentiable and continuous lower bounded functions respectively. Define $v_t(x, t) = \frac{\partial v(x,t)}{\partial t}$, $v_x(x, t) = [\frac{\partial v(x,t)}{\partial x_1}, \ldots, \frac{\partial v(x,t)}{\partial x_K}]$ and $v_{xx}(x, t) = [\frac{\partial^2 v(x,t)}{\partial x_i x_j}]_{K \times K}$. Under the conditions in (Feynman-Kac) Theorem 8.2.1 in Øksendal (2000) (or see Theorem 7.6 in Karatzas and Shreve 1991 for more general conditions), the function v is the unique solution to,

$$v_t(x_i, t_j) - v_x(x_i, t_j) b(x_i) - \frac{1}{2} tr\left[\sigma(x_i)^\mathsf{T} v_{xx}(x_i, t_j) \sigma(x_i) \right] + \varphi(x_i) v(x_i, t_j) = 0, \tag{41}$$

with initial condition $v(x, t_0) = g(x)$. The goal is to solve the partial differential equation in equation (41) for the unknown v and this can be done by numerical approximation using the Crank-Nicolson method. To simplify the exposition assume that $v : \mathbb{R} \times [t_0, t] \to \mathbb{R}$, thus the function v can be approximated on a set of discrete points in $I \times [t_0, t]$, with I denoting a compact set in $\mathbb{R}$. Let an arbitrary point in this

set be (x_i, t_j), with $i = 1, \cdots, N$, $j = 1, \cdots, J$, and a corresponding value of $V_{i,j}$. The finite number of points in $I \times [t_0, t]$ are equally spaced at distances Δx and Δt respectively. Consider de following approximations to construct the Crank-Nicolson algorithm,

$$v(x_i, t_j) \simeq \frac{V_{i,j} + V_{i,j-1}}{2}, \quad v_x(x_i, t_j) \simeq \frac{V_{i+1,j} + V_{i+1,j-1} - V_{i-1,j} - V_{i-1,j-1}}{4\Delta x}$$

$$v_t(x_i, t_j) \simeq \frac{V_{i,j} - V_{i,j-1}}{\Delta t}, \quad v_{xx}(x_i, t_j) \simeq \frac{V_{i+1,j} + V_{i+1,j-1} - 2(V_{i,j-1} + V_{i,j}) + V_{i-1,j} + V_{i-1,j-1}}{2(\Delta x)^2}.$$

The system of equations in equation (42) follows replacing the previous approximations in equation (41).

$$\underbrace{\begin{bmatrix} \eta_{1j}^* & -\kappa_{1j}^* & a_{1j} & 0 & \cdots & 0 & 0 & 0 & 0 \\ k_{2j} & \eta_{2j} & -\kappa_{2j} & 0 & \cdots & 0 & 0 & 0 & 0 \\ 0 & k_{3j} & \eta_{3j} & -\kappa_{3j} & \cdots & 0 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 0 & k_{N-1,j} & \eta_{N-1,j} & -\kappa_{N-1,j} \\ 0 & 0 & 0 & 0 & \cdots & 0 & c_{N,j} & k_{N,j}^* & \eta_{N,j}^* \end{bmatrix}}_{C_j} \underbrace{\begin{bmatrix} V_{1,j} \\ V_{2,j} \\ V_{3,j} \\ \vdots \\ V_{N-1,j} \\ V_{N,j} \end{bmatrix}}_{V_j} = \underbrace{\begin{bmatrix} d_{1,j}^* \\ d_{2,j} \\ \vdots \\ d_{N-1,j} \\ d_{N,j}^* \end{bmatrix}}_{d_j}$$

(42)

where $d_{i,j} \equiv \kappa_{ij} V_{i+1,j-1} + \mathfrak{n}_{ij} V_{i,j-1} - k_{ij} V_{i-1,j-1}$,

$$\kappa_{ij} = \frac{b(x_i)}{4\Delta x} + \frac{\sigma^2(x_i)}{4(\Delta x)^2}, \qquad\qquad \eta_{ij} = \frac{1}{\Delta t} + \frac{\sigma^2(x_i)}{2(\Delta x)^2} + \frac{\varphi(x_i)}{2}$$

$$k_{ij} = \frac{b(x_i)}{4\Delta x} - \frac{\sigma^2(x_i)}{4(\Delta x)^2}, \qquad\qquad \mathfrak{n}_{ij} = \frac{1}{\Delta t} - \frac{\sigma^2(x_i)}{2(\Delta x)^2} - \frac{\varphi(x_i)}{2}.$$

In this case where the drift and the diffusion coefficients do not depend on time the matric C_j does not depend on t either, thus $C_j = C$. The coefficients in the lower and upper rows of C_j are corrected due to the missing terms in V_j as follows

$$\eta_{1j}^* = \frac{1}{\Delta t} + \frac{b(x_1)}{4\Delta x} - \frac{\sigma^2(x_1)}{4(\Delta x)^2} + \frac{\varphi(x_1)}{2}, \qquad \kappa_{1j}^* = -\frac{\sigma^2(x_1)}{2(\Delta x)^2}$$

$$a_{1j} = -\kappa_{1j}, \qquad\qquad \mathfrak{h}_{1j} = \frac{1}{\Delta t} - \frac{b(x_1)}{4\Delta x} + \frac{\sigma^2(x_1)}{4(\Delta x)^2} - \frac{\varphi(x_1)}{2}$$

$$c_{N,j} = k_{Nj}, \qquad\qquad k_{N,j}^* = \frac{\sigma^2(x_N)}{2(\Delta x)^2}$$

$$\eta_{N,j}^* = \frac{1}{\Delta t} - \frac{b(x_N)}{4\Delta x} - \frac{\sigma^2(x_N)}{4(\Delta x)^2} + \frac{\varphi(x_N)}{2}, \qquad \mathfrak{l}_{N,j} = \frac{1}{\Delta t} + \frac{b(x_N)}{4\Delta x} + \frac{\sigma^2(x_N)}{4(\Delta x)^2} - \frac{\varphi(x_N)}{2}$$

with $d_{1j}^* \equiv \mathfrak{h}_{1j} V_{1,j-1} + \kappa_{1j}^* V_{2,j-1} - a_{1j} V_{3,j-1}$ and $d_{N,j}^* \equiv \mathfrak{l}_{N,j} V_{N,j-1} - k_{N,j}^* V_{N-1,j-1} - c_{N,j} V_{N-2,j-1}$.

As a result, V_j can be computed for any given time t_j solving the system in equation (42) recursively for all j. Note that $V_{i,0} = g(x_i)$ and consequently d_1 is known, so for all t_j

$$V_j = C_j^{-1} d_j. \tag{43}$$

Appendix E   A stochastic Runge-Kutta scheme

So far the proposed model accepts arbitrary (stable) Ito stochastic differential equations to model the intensity process and a suitable method to solve any of them numerically is the Runge-Kutta Method which converges strongly with order 1.[44] Consider a stochastic process X_t driven by

$$dX_t = b(X_t)\, dt + \sigma(X_t)\, dW_t, \text{ with } X_{t_0} = x_{t_0} \tag{44}$$

To compute an approximate solution on the interval $[t_0, t_J]$, construct a grid of points equally spaced at distances $\Delta t = \frac{t_J - t_0}{J}$, let ϵ_t be distributed as an standard normal and $\Delta W_{t_k} = W_{t_k + \Delta t} - W_{t_k}$, then each computed solution path x_t must satisfy

$$
\begin{aligned}
x_{t_k + \Delta t} &= x_{t_k} + b(x_{t_k})\, \Delta t + \sigma(x_{t_k})\, \Delta W_{t_k} \\
&\quad + \frac{1}{2}\sigma(x_{t_k})\left[\frac{\sigma\left(x_{t_k} + \sigma(x_{t_k})\sqrt{\Delta t}\right) - \sigma(x_{t_k})}{\sigma(x_{t_k})\sqrt{\Delta t}} \right] \left((\Delta W_{t_k})^2 - \Delta t \right) \\
&= x_{t_k} + b(x_{t_k})\, \Delta t + \sigma(x_{t_k})\sqrt{\Delta t}\, \epsilon_{t_k} \\
&\quad + \frac{1}{2}\left[\sigma\left(x_{t_k} + b(x_{t_k})\, \Delta t + \sigma(x_{t_k})\sqrt{\Delta t}\right) - \sigma(x_{t_k}) \right] \left(\epsilon_{t_k}^2 - 1 \right)\sqrt{\Delta t}
\end{aligned}
$$

with $X_{t_0} = x_{t_0}$.

An alternative version of the algorithm includes the term $b(x_{t_k})\, \Delta t$ in the approximated forward point to evaluate the discretized derivative (see for instance Burrage and Platen 1994), but the term can be omitted without losing the strong order of convergence (pag. 153 in Kloeden, Platen, and Schurz 1994).

---

[44]See Sauer (2013) for a definition of strong convergence of SDE solvers and Kloeden and Platen (1992) for further details on stochastic Runge-Kutta schemes.

References

AHN, D. H., AND B. GAO (1999): "A parametric nonlinear model of term structure dynamics," *Review of Financial Studies*, 12(4), 721–762.

AÏT-SAHALIA, Y. (2002): "Telling From Discrete Data Whether the Underlying ContinuousTime Model is a Diffusion," *Journal of Finance*, pp. 2075–2112.

AÏT-SAHALIA, Y. (2008): "Closed-Form Likelihood Expansions for Multivariate Diffusions," *The Annals of Statistics*, 36(2), 906–937.

AÏT-SAHALIA, Y., AND R. L. KIMMEL (2007): "Maximum Likelihood Estimation of Stochastic Volatility Models," *Journal of Financial Economics*, 83(2), 413–452.

BERNDT, A., R. DOUGLAS, D. DUFFIE, M. FERGUSON, AND D. SCHRANZ (2008): "Measuring default risk premia from default swap rates and EDFs," *Tepper School of Business*, p. 49.

BURRAGE, K., AND E. PLATEN (1994): "Runge-Kutta methods for stochastic differential equations," *Ann. Numer. Math*, 1(1-4), 63–78.

CHERNOZHUKOV, V., AND H. HONG (2003): "An MCMC approach to classical estimation," *Journal of Econometrics*, 115(2), 293–346.

DUFFIE, D., L. H. PEDERSEN, AND K. J. SINGLETON (2003): "Modeling sovereign yield spreads: A case study of Russian debt," *The Journal of Finance*, 58(1), 119–159.

DUFFIE, D., AND K. J. SINGLETON (1993): "Simulated Moments Estimation of Markov Models of Asset Prices," *Econometrica*, 61(4), 929–952.

DUFFIE, D., AND K. J. SINGLETON (1999): "Modeling term structures of defaultable bonds," *Review of Financial Studies*, 12(4), 687–720.

DUFFIE, D., AND K. J. SINGLETON (2003): *Credit risk: pricing, measurement, and management*. Princeton University Press, Princeton, N.J.

ENGLE, R. F., AND K. F. KRONER (1995): "Multivariate Simultaneous Generalized Arch," *Econometric Theory*, 11(1), 122–150.

FENTON, V. M., AND A. R. GALLANT (1996): "Convergence Rates of SNP Density Estimators," *Econometrica*, 64(3), 719–727, bibtex: Fenton1996.

GALLANT, A. R., AND J. R. LONG (1997): "Estimating stochastic diferential equations efficiently by minimum chi-squared," *Biometrika*, 84(1), 125–141.

GALLANT, A. R., AND D. W. NYCHKA (1987): "Semi-Nonparametric Maximum Likelihood Estimation," *Econometrica*, 55(2), 363–390.

GALLANT, A. R., AND G. TAUCHEN (1996): "Which Moments to Match?," *Econometric Theory*, 12(4), 657–681.

——— (2013a): "EMM: A program for efficient method of moments estimation, Version 1.6, User's Guide," *Unpublished discussion paper, Duke University, Durham, NC.*

——— (2013b): "SNP: A Program for Nonparametric Time Series Analysis, Version 9.0, User's Guide," *Manuscript, Duke University, Durham, NC.*

GALLANT, A. R., AND G. E. TAUCHEN (1988): *On fitting a recalcitrant series: the pound/dollar exchange rate, 1974-83.* H.G.B. Alexander Foundation, Chicago.

GARCIA, R., E. GHYSELS, AND E. RENAULT (2010): "The Econometrics of Option Pricing," in *Handbook of Financial Econometrics: Tools and Techniques*, ed. by Y. A.-S. P. HANSEN, vol. 1 of *Handbooks in Finance*, pp. 479–552. North-Holland, San Diego, Chapter 9.

GOURIEROUX, C., A. MONFORT, AND E. RENAULT (1993): "Indirect inference," *Journal of Applied Econometrics*, 8(S1), S85–S118.

HURN, A. S., J. I. JEISMAN, AND K. A. LINDSAY (2007): "Seeing the wood for the trees: A critical evaluation of methods to estimate the parameters of stochastic differential equations," *Journal of Financial Econometrics*, 5(3), 390–455.

HURN, A. S., K. A. LINDSAY, AND A. J. MCCLELLAND (2013): "A quasi-maximum likelihood method for estimating the parameters of multivariate diffusions," *Journal of Econometrics*, 172(1), 106–126.

JARROW, R. A. (2004): "Risky coupon bonds as a portfolio of zero-coupon bonds," *Finance Research Letters*, 1(2), 100–105.

JARROW, R. A., AND S. M. TURNBULL (1995): "Pricing Derivatives on Financial Securities Subject to Credit Risk," *The Journal of Finance*, 50(1), 53–85.

JENSEN, B., AND R. POULSEN (2002): "Transition densities of diffusion processes: numerical comparison of approximation techniques," *The Journal of Derivatives*, 9(4), 18–32.

JOHANNES, M. S., N. G. POLSON, AND J. R. STROUD (2009): "Optimal Filtering of Jump Diffusions: Extracting Latent States from Asset Prices," *Review of Financial Studies*, 22(7), 2759–2799.

KARATZAS, I., AND S. E. SHREVE (1991): *Brownian motion and stochastic calculus.* Springer-Verlag, New York.

KLOEDEN, P. E., AND E. PLATEN (1992): *Numerical solution of stochastic differential equations.* Springer-Verlag, Berlin.

KLOEDEN, P. E., E. PLATEN, AND H. SCHURZ (1994): *Numerical Solution of SDE Through Computer Experiments.* Springer Science & Business Media.

LANDO, D. (1998): "On cox processes and credit risky securities," *Review of Derivatives Research*, 2(2-3), 99–120.

LEE, B.-S., AND B. F. INGRAM (1991): "Simulation estimation of time-series models," *Journal of Econometrics*, 47(2–3), 197–205.

LONGSTAFF, F. A., S. MITHAL, AND E. NEIS (2005): "Corporate Yield Spreads: Default Risk or Liquidity? New Evidence from the Credit Default Swap Market," *The Journal of Finance*, 60(5), 2213–2253.

ØKSENDAL, B. (2000): *Stochastic Differential Equations: An Introduction with Applications*. Springer, Berlin ; New York, 5th edition edn.

PAKES, A., AND D. POLLARD (1989): "Simulation and the Asymptotics of Optimization Estimators," *Econometrica*, 57(5), 1027–1057.

PAN, J. (2002): "The jump-risk premia implicit in options: evidence from an integrated time-series study," *Journal of Financial Economics*, 63(1), 3–50.

PAN, J., AND K. J. SINGLETON (2008): "Default and Recovery Implicit in the Term Structure of Sovereign CDS Spreads," *The Journal of Finance*, 63(5), 2345–2384.

PASTORELLO, S., V. PATILEA, AND E. RENAULT (2003): "Iterative and Recursive Estimation in Structural Nonadaptive Models," *Journal of Business & Economic Statistics*, 21(4), 449–482.

PEDERSEN, A. R. (1995): "A New Approach to Maximum Likelihood Estimation for Stochastic Differential Equations Based on Discrete Observations," *Scandinavian Journal of Statistics*, 22(1), 55–71.

ROGERS, L. C. G., AND D. WILLIAMS (2000): *Diffusions, Markov Processes and Martingales: Volume 2, Itô Calculus*, vol. 2. Cambridge University Press, Cambridge, U.K. ; New York, 2 edition edn.

SAUER, T. (2013): "Computational solution of stochastic differential equations," *Wiley Interdisciplinary Reviews: Computational Statistics*, 5(5), 362–371.

SCHWARZ, G. (1978): "Estimating the Dimension of a Model," *The Annals of Statistics*, 6(2), 461–464.

SMITH, A. A. (1993): "Estimating nonlinear time-series models using simulated vector autoregressions," *Journal of Applied Econometrics*, 8(S1), S63–S84.

STRAMER, O., AND J. YAN (2007): "On Simulated Likelihood of Discretely Observed Diffusion Processes and Comparison to Closed-Form Approximation," *Journal of Computational and Graphical Statistics*, 16(3), 672–691.

STROOCK, D. W., AND S. R. S. VARADHAN (2005): *Multidimensional Diffusion Processes*. Springer, Berlin ; New York.

ZHANG, F. X. (2004): "What Did the Credit Market Expect of Argentine Default? Evidence from Default Swap Data," *No. 2003-25. Board of Governors of the Federal Reserve System (US)*.