Path dependent Modeling In Finance

Josef Teichmann

ETH Zürich

AQFC 2025, Shenzhen

Introduction

Goal of this talk is ...

- to present some classical results around Takens' theorem due to Floris Takens, Jaroslov Stark, and many others.
- to present a continuous time version of Takens' theorem.
- to highlight on some applications to (financial) time series analysis..

Introduction

Goal of this talk is ...

- to present some classical results around Takens' theorem due to Floris Takens, Jaroslov Stark, and many others.
- to present a continuous time version of Takens' theorem.
- to highlight on some applications to (financial) time series analysis..

Introduction

Goal of this talk is ...

- to present some classical results around Takens' theorem due to Floris Takens, Jaroslov Stark, and many others.
- to present a continuous time version of Takens' theorem.
- to highlight on some applications to (financial) time series analysis..

Classical Modeling

It is a classical modeling paradigm to specify a state space M, on which a certain (Markovian) dynamics takes place, and to specify a transition operator $T : M \to M$, which specifies the dynamics locally.

The evolution according to the dynamics T is then given by $(T^k(x))_{k \geq 0}$.

Often we do not observe the state space itself, but rather $\varphi : M \to \mathbb{R}$, e.g. we collect knowledge on $(\varphi(T^k(x)))_{k \geq 0}$. We call this sequence (time series) the *observations*.

It is an important question in which relation the observations are to the original dynamics. Floris Takens' theorem from 1980 gives a very surprising answer to this question.

Informal statement of Takens' theorem

Let M be a m dimensional compact manifold. Let T and φ be generic (and satisfying some mild regularity conditions), then the map

$$\Phi := \Phi_{(T,\varphi)} : x \mapsto (\varphi(T^k x))_{k=0,\ldots,2m}$$

embeds M into $\mathbb{R}^{2m+1}$. This means in particular that

$$\varphi(T^{2m+1}x) = G(\varphi(x), \varphi(Tx), \ldots, \varphi(T^{2m}x))$$

for some function G on $\Phi(M)$. On $\Phi(M)$ the dynamics appears in a normalized form as shift to the left combined with G:

$$G := \pi_{2m} \circ \Phi \circ T \circ \Phi^{-1}$$

and is well defined (π_{2m} projects on the last component).

Consequences

It follows from this surprising result ...

- that the dynamics, expressed through $T : M \to M$ in a Markovian way can also be expressed on $\Phi(M)$ via G and a shift.
- that the state space M can be realized (in the sense of an embedding) as a subset of $\mathbb{R}^{2m+1}$ (numerical model).
- that we can write a delay equation instead of the Markovian representation, i.e.

$$y = (y_0, \ldots, y_{2m}) \in \Phi(M) \mapsto (y_1, \ldots, y_{2m}, G(y)) \in \Phi(M)$$

  is an equivalent dynamics of delay type. This has tremendous consequences in learning theory (just learn G to represent a dynamics numerically without constructing the state space neither the full dynamics).
- that results with multivariate observation $\varphi : M \to \mathbb{R}^l$ are formulated accordingly. Important is that the embedding space is more than $2m$ dimensional.

Consequences

It follows from this surprising result ...

- that the dynamics, expressed through $T : M \to M$ in a Markovian way can also be expressed on $\Phi(M)$ via G and a shift.
- that the state space M can be realized (in the sense of an embedding) as a subset of $\mathbb{R}^{2m+1}$ (numerical model).
- that we can write a delay equation instead of the Markovian representation, i.e.

$$y = (y_0, \ldots, y_{2m}) \in \Phi(M) \mapsto (y_1, \ldots, y_{2m}, G(y)) \in \Phi(M)$$

  is an equivalent dynamics of delay type. This has tremendous consequences in learning theory (just learn G to represent a dynamics numerically without constructing the state space neither the full dynamics).
- that results with multivariate observation $\varphi : M \to \mathbb{R}^l$ are formulated accordingly. Important is that the embedding space is more than $2m$ dimensional.

Consequences

It follows from this surprising result ...

- that the dynamics, expressed through $T : M \to M$ in a Markovian way can also be expressed on $\Phi(M)$ via G and a shift.
- that the state space M can be realized (in the sense of an embedding) as a subset of $\mathbb{R}^{2m+1}$ (numerical model).
- that we can write a delay equation instead of the Markovian representation, i.e.

$$y = (y_0, \ldots, y_{2m}) \in \Phi(M) \mapsto (y_1, \ldots, y_{2m}, G(y)) \in \Phi(M)$$

  is an equivalent dynamics of delay type. This has tremendous consequences in learning theory (just learn G to represent a dynamics numerically without constructing the state space neither the full dynamics).
- that results with multivariate observation $\varphi : M \to \mathbb{R}^l$ are formulated accordingly. Important is that the embedding space is more than $2m$ dimensional.

Consequences

It follows from this surprising result ...

- that the dynamics, expressed through $T : M \to M$ in a Markovian way can also be expressed on $\Phi(M)$ via G and a shift.
- that the state space M can be realized (in the sense of an embedding) as a subset of $\mathbb{R}^{2m+1}$ (numerical model).
- that we can write a delay equation instead of the Markovian representation, i.e.

  $$y = (y_0, \dots, y_{2m}) \in \Phi(M) \mapsto (y_1, \dots, y_{2m}, G(y)) \in \Phi(M)$$

  is an equivalent dynamics of delay type. This has tremendous consequences in learning theory (just learn G to represent a dynamics numerically without constructing the state space neither the full dynamics).
- that results with multivariate observation $\varphi : M \to \mathbb{R}^l$ are formulated accordingly. Important is that the embedding space is more than $2m$ dimensional.

Delay equation models versus Markov models

In the landscape of models we can distinguish between delay equation models and Markovian models. Classically Markovian models are very popular but delay equation models are also successful, in particular recently.

Well known examples of delay equation modeling include ...

- joint models for S&P and VIX indices of delay equation type, Volterra equations, rough volatility models, etc.
- delay embedding models for classical dynamical systems.
- large language models.

Delay equation models versus Markov models

In the landscape of models we can distinguish between delay equation models and Markovian models. Classically Markovian models are very popular but delay equation models are also successful, in particular recently.

Well known examples of delay equation modeling include ...

- joint models for S&P and VIX indices of delay equation type, Volterra equations, rough volatility models, etc.
- delay embedding models for classical dynamical systems.
- large language models.

Delay equation models versus Markov models

In the landscape of models we can distinguish between delay equation models and Markovian models. Classically Markovian models are very popular but delay equation models are also successful, in particular recently.

Well known examples of delay equation modeling include ...

- joint models for S&P and VIX indices of delay equation type, Volterra equations, rough volatility models, etc.
- delay embedding models for classical dynamical systems.
- large language models.

Classical version [Takens 1980]

Let M be a compact manifold of dimension m and $T : M \to M$ be a smooth diffeomorphism. There exists an open dense set of $(T, \varphi) \in D^\infty(M, M) \times C^\infty(M, \mathbb{R})$ (referred to later as *generic*) such that

$$\Phi_{(T,\varphi)} : M \to \mathbb{R}^{2m+1}$$

is an embedding.

(We do not achieve minimal regularity assumptions here.)

Remarks

- Of course several improvements, in particular of the genericity condition, are possible and done.
- We aim, however, for a controlled version of the result, i.e. a result for systems of the type

$$x_{k+1} = T(x_k, y_k)\,;\; y_{k+1} = g(y_k)\,;\; x_0 = x \in M \text{ and } y_0 = y \in N\,.$$

  Notice that we know the forcing dynamics g, but we do *not* observe y.

Remarks

- Of course several improvements, in particular of the genericity condition, are possible and done.
- We aim, however, for a controlled version of the result, i.e. a result for systems of the type

$$x_{k+1} = T(x_k, y_k)\,;\; y_{k+1} = g(y_k)\,;\; x_0 = x \in M \text{ and } y_0 = y \in N\,.$$

  Notice that we know the forcing dynamics g, but we do *not* observe y.

Classical version for controlled systems [Stark 1999]

Let M be a compact manifold of dimension m, N be a compact manifold of dimension n and $T : M \times N \to M$ a smooth map such that $T_y; = T(., y)$ is a smooth diffeomorphism for each $y \in N$. Let furthermore g be a diffeomorphism. Assume that the periodic orbits of g of period less than $2d$ are distinct and $d > 2(m+n)+1$, then there exists a generic set of (T, φ) such that

$$(x, y) \mapsto \Phi_{(T,g,\varphi)}(x, y) := (\varphi(x), \varphi(T(x, y)), \ldots, \varphi(T(x_{d-1}, y_{d-1}))) \in \mathbb{R}^d$$

is an embedding.

Classical version for stochastic systems [Stark 2003]

Let M be a compact manifold of dimension m, N be a compact manifold of dimension n and $T : M \times N \to M$ a smooth map such that $T_y; = T(., y)$ is a smooth diffeomorphism for each $y \in N$. Let furthermore μ be a measure on N absolutely continuous with respect to Lebesgue measure. Let $d > 2m$, then there exists a generic set of (T, φ) such that

$$x \mapsto \Phi_{(T,g,\varphi)}(x) := (\varphi(x), \varphi(T(x_0, \omega_0)), \ldots, \varphi(T(x_{d-1}, \omega_{d-1}))) \in \mathbb{R}^d$$

is an embedding almost surely with respect to μ^∞.

(Notice that we have to know the stochastic driver $\omega = (\omega_k)_{k \geq 0}$ to construct the random embedding.)

Remarks

- The random embedding leads to a delay equation representation of type

$$G(\varphi(x), \varphi(T(x_0, \omega_0)), \ldots, \varphi(T(x_{d-1}, \omega_{d-1})), \omega_0, \ldots, \omega_{d-1}, \omega_d),$$

  to represent $\varphi(T(x_d, \omega_d))$, which hold almost surely with respect to μ^∞.
- Generalizations for the type of measure on N^∞ are of course possible and known.

Remarks

- The random embedding leads to a delay equation representation of type

$$G(\varphi(x), \varphi(T(x_0, \omega_0)), \ldots, \varphi(T(x_{d-1}, \omega_{d-1})), \omega_0, \ldots, \omega_{d-1}, \omega_d),$$

  to represent $\varphi(T(x_d, \omega_d))$, which hold almost surely with respect to μ^∞.
- Generalizations for the type of measure on N^∞ are of course possible and known.

Stochastic Setting for this talk

Let X be the solution of the Markovian system on some high dimensional compact state space M, which, for simplicity, is embedded in $\mathbb{R}^n$ (and allows us to use Ito calculus without further ado).

$$dX_t^x = \sum_{i=0}^{r} V^i(X_t) dB_t^i \, ; \; X_0^x = x$$

with smooth vector fields $V^0, \ldots, V^r$ of C_b^∞ type well defined around M.

We shall now reformulate deterministic and stochastic Takens in this setting (analogously the controlled version could be reformulated).

The deterministic case $r = 0$

For any $T > 0$ the map

$$x \mapsto \Phi_{(V,\varphi)}(x) := \big(t \mapsto \varphi(X_t^x)\big)_{t\in[0,T]} \in C([0,T],\mathbb{R})$$

is an embedding for a generic set of (V^0, φ).

(Notice the absence of dimension of M in this formulation, but the infinite amount of information instead.)

The stochastic case $r > 0$

For any $T > 0$ the map

$$x \mapsto \Phi_{(V,\varphi)}(x) := \big(t \mapsto \varphi(X_t^x)\big)_{t \in [0,T]} \in C([0,T], \mathbb{R})$$

is an embedding for a generic set of $(V^0, \dots, V^r, \varphi)$ almost surely with respect to Wiener measure.

(Notice again the absence of dimension of M in this formulation, but the infinite amount of information instead.)

Interpretation for $r = 0$

Since

$$\frac{d}{ds}|_{s=0}\varphi(X^x_{t+s}) = d\varphi(X^x_t)V(X^x_t)$$

we obtain a representation as delay differential equation

$$d_s\varphi(X^x_s) = G(\varphi(X^x_{[s-T,s]})ds$$

Interpretation for $r > 0$

Completely analogously we obtain a representation as delay stochastic differential equation

$$d_s\varphi(X_s^x) = \sum_{i=0}^{r} G^i(\varphi(X_{[s-T,s]}^x)dB_s^i$$

where, however, G^i depends on the stochastic driver B (continuity in rough path sense with respect to the driving noise can be achieved, but certainly regularity with respect to the observation path).

An example and an (easy) stochastic proof

Given the above Ito diffusion on a compact state space, where we observe the d Brownian motions together with $(\varphi(X_t^x))_{t \geq 0}$, then an iteration of

$$(\varphi(X^x), B^i) \mapsto \frac{d}{dt}[\varphi(X^x), B^i] = d\varphi(X^x) V^i(X^x)$$

constructs (in the generic case) sufficiently many features, which become point separating on M, whence yielding an embedding.

It shows that we can (in case of high frequency observations of the process together with the noise) construct the embedding.

Theorem

If one can reconstruct from observations the Brownian motion (or a more general stochastic process driving the evolution), then a representation of $(\varphi(X_t^x))_{t\geq 0}$ as stochastic delay equation with regular characteristics with respect to noise and observation path is possible.

A reconstruction is often possible via the quadratic covariation: let $Y_t := \psi(X_t^x)$ for a sufficiently regular other observation map ψ such that

$$dY_t = \mu_t dt + \Sigma_t dB_t$$

with almost surely invertible Σ_t (this is an assumption!). Then under mild conditions

$$B + \int_0^{\cdot} \Sigma_t^{-1}\mu_t dt = \Big(\sqrt{\frac{d}{dt}[Y,Y]}^{-1} \bullet Y\Big)\,.$$

With an estimation of the drift parameters one obtains a fully specified stochastic delay equation representation for $\varphi(X^x)$.

An example

Consider a (unknown) high dimensional Ito dynamics $(X_t^x)_{t\geq 0}$, where $(\varphi(X_t^x))_{t\geq 0}$ describes a continuously observed time series of discounted price processes. A subset of those prices is enough to identify B up to the market price of risk. Under the equivalent martingale measure we obtain a delay representation of the dynamics, which can be learned in terms of the strong representation

$$\varphi(X_{t+\Delta}^x) - \varphi(X_t^x) = \sum_w I^w\big((\varphi(X_{t-s}^x)_{[-T,0]}\big)\,\mathrm{Sig}_{t,t+\Delta}^w(B)$$

which holds almost surely with respect to the *equivalent martingale measure*. I^w as a function of the path can be learned efficiently from the observations. This is a regression based non-linear time series model.

Background UAT: integrable maps on path spaces can be strongly approximated by, e.g., signature components (see work with Christa Cuchiero and Philipp Schmocker on weighted spaces).

References

- Jaroslav Stark, Delay Embeddings for Forced Systems, Part I: Deterministic Forcing, J. Non-linear Sci., 9, 255–332, 1999.
- Jaroslav Stark, David Broomhead, Mike Davies, J. Huke, Delay Embeddings for Forced Systems, Part II: Stochastic Forcing, J. Non-linear Sci., 13, 519–577, 2003.
- Floris Takens, Detecting Strange Attractors in Fluid Turbulence. In D.A. Rand and L.-S. Young, eds., Dynamical Systems and Turbulence, Warwick 1980.