

Gaussian Agency problems with memory and Linear Contracts

Eduardo Abi Jaber, Stéphane Villeneuve

▶ **To cite this version:**

Eduardo Abi Jaber, Stéphane Villeneuve. Gaussian Agency problems with memory and Linear Contracts. Finance and Stochastics, 2022. hal-03783062

HAL Id: hal-03783062
https://hal.science/hal-03783062

Submitted on 21 Sep 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Gaussian Agency problems with memory and Linear Contracts

Eduardo Abi Jaber*
Stéphane Villeneuve†

September 21, 2022

Abstract

Can a principal still offer optimal dynamic contracts that are linear in end-of-period outcomes when the agent controls a process that exhibits memory? We provide a positive answer by considering a general Gaussian setting where the output dynamics are not necessarily semi-martingales or Markov processes. We introduce a rich class of principal-agent models that encompasses dynamic agency models with memory. From the mathematical point of view, we develop a methodology to deal with the possible non-Markovianity and non-semimartingality of the control problem, which can no longer be directly solved by means of the usual Hamilton-Jacobi-Bellman equation. Our main contribution is to show that, for one-dimensional models, this setting always allows for optimal linear contracts in end-of-period observable outcomes with a deterministic optimal level of effort. In higher dimension, we show that linear contracts are still optimal when the effort cost function is radial and we quantify the gap between linear contracts and optimal contracts for more general quadratic costs of efforts.

Keywords Principal-Agent models, Continuous-time control problems.

1 Introduction

The extensive literature analyzing the dynamic principal-agent problem has shown that it is important but difficult to design the optimal shape of contracts in a tractable way. Indeed,

*CMAP, Ecole Polytechnique Paris, route de Saclay 91128 Palaiseau Cedex

†Université Toulouse 1 Capitole (TSE-TSM-R), 1 esplanade de l'université, 31000 Toulouse, stephane.villeneuve@tse-fr.eu. The author acknowledges funding from ANR PACMAN and from ANR under grant ANR-17-EUR-0010 (Investissements d'Avenir program) and gratefully thanks the FdR-SCOR "Chaire Marché des risques et création de valeurs". We are grateful to Thomas Mariotti and Takuro Yamashita for their comments and supports. We thank participants of Financial and Actuarial Mathematics Seminar at Liverpool, Financial and Insurance Mathematics at ETH,Math. Finance Seminar at Bielefeld, 10th AMaMeF Conference at Padova, Annual TSE-SCOR conference at Toulouse for their helpful comments and suggestions.

optimal contracts in dynamic agency problems are generally defined as complex functionals of a stream of contractible variables, such as revenues. Moreover, as first identified by Rogerson (1985) (see also Laffont and Martimort (2009)) theoretical contracts exhibit memory, even in the most commonly used models that assume uncorrelated shocks, which unfortunately prevents them from matching real-world practices (see Bolton and Dewatripont (2005)). In addition, firms' revenues empirically show long memory[1] and we lack a theoretical framework that justifies the signing of simple tractable contracts in an environment with inter-temporal links across time periods.
In a Brownian setting, the breakthrough paper by Holmstrom and Milgrom (1987) (HM) shows that the optimal contract is linear in profits under some specific assumptions: the agent exerts effort continuously, principal and agent have CARA utilities, the agent bears a pecuniary cost of effort and finally the outcomes generated in the absence of effort are modeled by a fully observable Brownian motion. Since then, several attempts have been made to obtain closed-form contracts in environments that relax at least one of the assumptions of the HM model. Sung (1995) showed that the optimal contract is still linear when the agent also controls the variance of the output. Hellwig and Schmidt (2002) showed that linear contracts are nearly optimal in a discrete-time version of the HM model. Edmans and Gabaix (2011) and Edmans et al. (2012) obtain striking general results in a discrete-time model where none of the four hypotheses is retained but where the agent makes its decision in each period after having observed the noise. However, they focus primarily on the cheapest implementation of a particular action, rather than on the objective of maximizing the principal's preference. Another important recent contribution beyond the Holmstrom and Milgrom setting has been made by Carroll (2015) who showed that optimal contracts are linear in a general one-period model with uncertainty.

In this article, we enrich the Holmstrom and Milgrom modeling framework by going beyond the assumption that the revenues are driven by a Brownian motion.

We will instead consider Volterra Gaussian processes that are a generalization of the standard Brownian motion to study time-dependent effects. More precisely, a Volterra Gaussian process is a Wiener integral process with respect to a standard Brownian motion involving a deterministic integrand called -*kernel*. Thus, at every point in time, it is an infinite linear combination of i.i.d. Gaussian random variables with time-dependent coefficients. Although we are aware of the shortcomings of the three remaining HM assumptions[2], our targeted choice is primarily motivated by the fact that, by allowing arbitrary integrand kernel functions in the Wiener integral, our Volterra Gaussian agency model encompasses agency models with short and long run autocorrelations. In particular, one of the main examples of Volterra Gaussian processes is the mean-reverting process which

[1] In this paper, we use indifferently the terms long (short) memory and long (short)-range dependence see definition 2.1. page 42 in Beran (1994).

[2] For instance, Edmans and Gabaix (2011) clearly argue *there is ample evidence of decreasing absolute risk aversion, and many effort decisions do not involve a monetary expenditure*

allows us to get closer of recent models of dynamic contracts with persistence such as those developed in Williams (2011) or career concerns as in Cisternas (2018).

For a long time, Volterra Gaussian processes have been considered as a natural tool for modeling continuous phenomena with memory. In particular, the fractional Brownian motion (FBM), a Volterra Gaussian process with short and long range dependence, initially introduced by Kolmogorov (1940), was popularized by Mandelbrot and Van Ness (1968) in finance to model the empirically-validated long-term dependence of stock returns. More recently, a stream of literature suggested the use of variants of the fractional Brownian motion in stochastic volatility modeling to capture the roughness of the time series of the volatility of an asset which has been observed empirically in the market, see Gatheral et al. (2018); Abi Jaber et al. (2019) and the references therein. In general, such processes are non-Markovian and non-semimartingales, which make their study more intricate, both theoretically and practically, and prevents the use of standard stochastic calculus tools. Within the framework of optimal dynamic contracting theory, the continuous-time semi-martingale models have received a lot of attention the past thirty years, when a significant progress has been made by relying on the recursive approach pioneered by Green (1987), Spear and Srivastava (1987), Thomas and Worrall (1990). Discrete-time models were first developed (see Clementi and Hopenhayn (2006); DeMarzo and Fishman (2007), followed by Biais et al. (2007)), while the breakthrough paper by Sannikov (2008) resulted in the recent study of dynamic contracting in continuous-time models (see also DeMarzo and Sannikov (2006); Biais et al. (2010); DeMarzo et al. (2012)). The main advantage of continuous-time semi-martingale models lies in the fact the procedure to find the optimal contract can be embedded in the standard theory of Markovian stochastic control using the theory of martingales and stochastic calculus, see Schättler and Sung (1993) for a general presentation and Cvitanić et al. (2018) for a rigorous mathematical justification. Hence, under a fairly general set of assumptions, the contract can be characterized unambiguously by solving an Hamilton-Jacobi-Bellman equation where the so-called agent promised value plays the role of a state variable.

In allowing a non-semimartingale and non-Markovian setting, this paper makes an additional methodological contribution to solve for the optimal contract. The main idea is to use the so-called martingale optimality principle to study the agent and the principal problem sequentially as a Stackelberg game. The first step is to offer a class of incentive-compatible contracts by revisiting the martingale approach of Schättler and Sung (1993) and Sannikov (2008). The second step and our main contribution is to explicitly solve the principal problem which becomes a controlled stochastic Volterra problem. This requires the introduction of an auxiliary state variable - *the effort-corrected forward output* - which captures all the non-Markovianity and allows the application of the martingale optimality principle for the principal problem. In a one-dimensional setting, our key result is that the optimal contract is linear in the terminal value of the output, although the principal has in

general a coarser information than the agent. The optimal contract has interesting features. The slope or marginal value of the contract is independent of the output dynamics, only the intercept depends on the latter as a function of the optimal effort which is proportional to the kernel. Therefore, random parts of contracts signed in different one-dimensional Gaussian environments are identical although the required effort levels are environment specific, deterministic and exhibit interesting features in relation with the properties of the kernel.

We extend the paper with a discussion of the optimality of linear contracts in higher dimension. We address the multitask principal-agent problem in which a principal with CARA preferences hires a single agent with CARA preferences to perform different tasks. The outcome of each task is assumed to follow a Volterra Gaussian process whose evolution depends on the agent's continuous effort in each task while the profit is the sum of these different outcomes. Our main result is that there is no value in observing the agent's activities separately when the cost of effort is assumed to be a radial quadratic function. Under this assumption, the optimal contract only uses aggregate information and is still linear in the end-of-period outcome. When we consider a general effort cost function, we characterize the contract that would be optimal if the principal were able to observe the Brownian filtration and measure the utility gap when a less-informed principal forces herself to sign the best linear contract and identify factors that reduce the loss of utility associated with the use of linear contracts. This paper thus shows that linear contracts can closely achieve maximum principal utility in Gaussian environments.

2 The one-dimensional model

In this section, we present the economic model which is essentially an extension of the Holmstrom and Milgrom (1987) framework.

General Description: We consider a risk-averse investor, who owns a project and signs a fixed-term contract with a risk-averse manager, the latter being necessary to operate a project. Time is continuous and the time horizon is $T > 0$. In the absence of effort, the stochastic output process $(X_t)_{t\leq T}$ of the project evolves up to time T as

$$X_t = g_0(t) + \int_0^t K(t,s)dB_s, \tag{2.1}$$

where B is a standard one-dimensional Brownian motion, $g_0 : [0,T] \to \mathbb{R}$ is a measurable deterministic input function, $K : [0,T]^2 \to \mathbb{R}$ is a measurable Volterra Kernel, i.e. $K(t,s) = 0$ for $s \geq t$ such that

$$\sup_{t\leq T} \int_0^T K^2(t,s)ds < \infty.$$

We first observe that our model encompasses a large class of output dynamics that offers great modeling flexibility to model agency relationships in different sectors of the economy. Obviously, this setting contains the Holmstrom and Milgrom Brownian model by choosing $g_0(t) = x_0$ and $K(t,s) = \sigma$ for any pair $s < t$ for some constant σ. Even more generally, the case of a time-dependent volatility can be recovered by setting $K(t,s) = \sigma(s)\mathbb{1}_{s<t}$ for some square-integrable function σ. More interestingly, this framework also contains mean-reverting processes which are widely used to model output in the energy and mining sectors, if for instance, we choose $g_0(t) = e^{-\lambda t}x_0 + \frac{\mu_0}{\lambda}(1 - e^{-\lambda t})$ and $K(t,s) = e^{-\lambda(t-s)}\mathbb{1}_{s<t}$, the output then follows the Ornstein-Ulhenbeck dynamics

$$dX_t = (\mu_0 - \lambda X_t)\, dt + dB_t.$$

Another example is the Brownian bridge pinned for instance in 0 at some time $T_0 > T$ which falls into this category with a kernel given by

$$K(t,s) = \frac{T_0 - t}{T_0 - s}\mathbb{1}_{s<t}.$$

The Brownian bridge has the semi-martingale decomposition

$$dX_t = -\frac{X_t}{T_0 - t}\, dt + dB_t,$$

and may be used in any situation where the agent has access to information about the future output. For example, it can be used to model the output of a seasonal crop which will end up being zero after the harvest season. A more striking example is given by the family of fractional Gaussian processes such as:

- the Riemann-Liouville fractional Brownian motion where for $s < t$,

$$K(t,s) = c_H(t-s)^{H-1/2}, \quad H \in (0,1), \text{ for some constant } c_H.$$

- the fractional Brownian motion whose covariance function is $\Sigma_0(s,u) = \frac{1}{2}(s^{2H} + u^{2H} - |s-u|^{2H})$, for $H \in (0,1)$, admits the Volterra representation (2.1) with the kernel

$$K(t,s) = \mathbb{1}_{s<t}\frac{(t-s)^{H-1/2}}{\Gamma(H+\frac{1}{2})}\, {}_2F_1\left(H - \frac{1}{2}; \frac{1}{2} - H; H + \frac{1}{2}; 1 - \frac{t}{s}\right),$$

  where ${}_2F_1$ is the Gauss hypergeometric function, see Decreusefond and Ustunel (1999).

Both types of fractional Brownian motion do not fall into the semi-martingale and Markovian frameworks when the so-called Hurst parameter H is different than $1/2$ (which corresponds to the case of the standard Brownian motion): they exhibit long range dependence

when $H > 1/2$ and short run dependence when $H < 1/2$ and prove to be statistically very good models for industries related to power generation.
Even more remarkably, equations with delay in the drift in the form of linear integro-differential convolution equations:

$$dX_t = \left(h(t) + \int_{[0,t]} \mu(ds) X_{t-s} \right) dt + \sigma dB_t, \tag{2.2}$$

with initial condition $X_0 \in \mathbb{R}$, where $h : [0,T] \to \mathbb{R}$, $\mu : \mathcal{B}([0,T]) \to \mathbb{R}$ of bounded variation, $X_0 \in \mathbb{R}$ and $\sigma \in \mathbb{R}$ admit a unique solution in the form of a Gaussian Volterra process (2.1) for some specific choice of input curve $g_0 : [0,T] \to \mathbb{R}$ and convolution kernel $K : [0,T]^2 \to \mathbb{R}$, see Appendix 7.2 for a detailed presentation of this observation. For instance, setting $\mu(dt) = \sum_{k=1}^m a_k \delta_{t_k}$, we recover equations with delay. Such equations fall into the semi-martingale framework, but are clearly not Markovian.
While very different in nature and in modeling objective, all these dynamics have in common to be Gaussian processes.

Another important observation about this framework relates to assumptions about the asymmetry in information, which has been interpreted by Holmstrom and Milgrom (1987) as a distinction between linear optimal contracts in outcomes X and those in accounts B. We denote by $\mathbb{F}^B$ the augmented filtration generated by $(B_t)_{t \leq T}$ and $\mathbb{F}^X$ the one generated by the output process $(X_t)_{t \leq T}$. It readily follows from (2.1) that $\mathbb{F}^X \subset \mathbb{F}^B$. Hereafter, we assume that the agent has better information than the principal about the project in the sense that he has access to the full information $\mathbb{F}^B$ while the principal observes only some aggregated information generated by the output $\mathbb{F}^X$. In general, these two filtrations do not coincide even in one-dimensional models as shown in the following example corresponding to a situation where the principal observes the output in a discretionary way.

Discrete observations of a Brownian motion: Assume $X_t = f(t)B_t$ where f is a bounded function on $[0,T]$. Observe that X is a Volterra process with $K(t,s) = f(t)\mathbb{1}_{s \leq t}$. Consider a subdivision $0 < t_1 < \ldots < t_n = T$ of the interval $[0,T]$ and let f be the function defined as a linear combination of unit impulses

$$f(t) = \sum_{i=1}^N \mathbb{1}_{t_i}(t). \tag{2.3}$$

The output process is purely discontinuous with $X_{t_i} = B_{t_i}$ and $X_t = 0$ for $t \neq t_i$ and may correspond to a situation where the principal performs audits at regular intervals. Therefore, $\mathbb{F}^X$ is strictly included in $\mathbb{F}^B$. We deduce that, even in a situation where the principal knows the agent is not exerting effort, the principal has a coarser information than the agent. In Volterra Gaussian models, we must therefore be careful that there may be asymmetric information between the principal and the agent regardless of the agency

problem we introduce below.

Agency problem: We assume that the agent can exert a continuous effort $(a_t)_{t\leq T}$ that modifies the probability distribution of X as follows

$$X_t = g_0(t) + \int_0^t K(t,s)(a_s\, ds + dB_s^a), \tag{2.4}$$

where B^a is also a Brownian motion. As is customary in the agency theory literature, while the output process X is observable by both players, the effort is the agent's private information. The agent's cost for exercising some effort level a is modeled through a strictly convex C^2 function $k(a)$ satisfying $k(0) = 0$. To alleviate the exposition, we will assume hereafter that the effort cost function is quadratic,

$$k(a) = \kappa\frac{a^2}{2}, \quad \text{for some } \kappa > 0. \tag{2.5}$$

Hereafter and in accordance with the paper of Holmstrom and Milgrom (1987), we model the preferences of the principal and the agent with CARA utility functions that are given respectively by

$$U_P(x) := -\exp(-\gamma_P x) \text{ and } U_A(x) := -\exp(-\gamma_A x), \quad \forall x \in \mathbb{R}.$$

In the beginning of the relationship, principal and agent agree on a contract of maturity T. To foster incentives, the contrat specifies a payment at time T which is modeled by a random variable ξ that is supposed to be $\mathcal{F}_T^X$ measurable. We assume that both players can fully commit to the contract and that the agent has a reservation utility level $R_0 = U_A(y_0) < 0$ below which he will refuse the contract. The latter inequality is referred to the participation constraint of the agent who has the option to reject a contract and enjoy a utility of autarky R_0.

Description of the probabilistic background: For completeness, we recall the rigorous formulation of the agency problem in order to make understandable the first-order conditions that we will give in the next section. Let $(\Omega, \mathcal{F}, \mathbb{F} := (\mathcal{F}_t)_{t\leq T}, \mathbb{P}_0)$ be a filtered probability space on which a $\mathbb{F}$–Brownian motion $B := (B_t)_{t\leq T}$ is defined with natural (completed) filtration $\mathbb{F}^B := (\mathcal{F}_t^B)_{t\leq T}$.

The firm's output or cash-flows observed by the principal are given by a stochastic process X with dynamics under $\mathbb{P}_0$,

$$X_t = g_0(t) + \int_0^t K(t,s)dB_s,$$

The impact of the agent's effort is modeled as a change of probability measure which changes the drift of the driving Brownian process. More precisely, agent's admissible actions are given by the following set

$$\mathcal{A} = \left\{ (a_t)_{t\leq T} \, \mathbb{F}\text{-progressively measurable: there exists } A > 0 \text{ s.t. } \int_0^T a_s^2 ds \leq A, \ \mathbb{P}_0 - \text{a.s.} \right\}.$$

Observe that the set of admissible actions $\mathcal{A}$ is not empty because it contains bounded actions. Clearly, any admissible process $a \in \mathcal{A}$ satisfies the Novikov's criterion

$$\mathbb{E}\left[\exp\left(\frac{1}{2}\int_0^T a_s^2 ds\right)\right] < +\infty,$$

which ensures that the process $\left(\exp\left(\int_0^T a_s \, dB_s - \frac{1}{2}\int_0^T a_s^2 \, ds\right)\right)_{0\leq t\leq T}$ is a martingale, see Karatzas and Shreve (1991, Proposition 5.12 p. 198). We can therefore define a family of equivalent probability measures $\mathbb{P}_a$ by

$$\frac{d\mathbb{P}_a}{d\mathbb{P}_0} = \exp\left(\int_0^T a_s \, dB_s - \frac{1}{2}\int_0^T a_s^2 \, ds\right),$$

where a ranges trough $\mathcal{A}$. Under $\mathbb{P}_a$, the process $B^a = B - \int_0^{\cdot} a_s \, ds$ is a $\mathbb{F}-$Brownian motion by Girsanov theorem and X evolves as

$$X_t = g_0(t) + \int_0^t K(t,s)(a_s \, ds + dB_s^a). \tag{2.6}$$

Because, the effort is unobservable, the principal only observes the trajectory of the output process X, the deterministic curve g_0 but not the last two terms of the decomposition (2.6) separately.
Interestingly, in the case of general Volterra processes, this model leads to a novel simple setting where we have persistence of past efforts on the output variation. To understand this, let us imagine that the agent makes a constant effort a on the interval $[0,t]$ and then stops exerting effort after t. Then, we have for $h > 0$, that

$$\mathbb{E}[(X_{t+h} - X_t)|\mathcal{F}_t] = g_0(t+h) - g_0(t) + \int_0^t (K(t+h,s) - K(t,s))(dB_s^a + a \, ds),$$

which induces persistence of past efforts on the future output increments whenever the functions $K(t+h,.)$ and $K(t,.)$ are not identical. Notice that in the HM model, the kernel is constant, so we recover that past efforts have no influence on future variations of the output.

The Principal-agent problem: It is well known that principal-agent relationships can be viewed as a Stackelberg game. The principal moves first by offering a contract that consists

in a compensation ξ, which belongs to the set of $\mathcal{F}_T^X$ measurable random variables, to the agent. The latter then reacts by choosing an effort policy based on the information available at each date inducing a probability measure $\mathbb{P}_a$. For any given contract ξ, let $V_0^A(\xi)$ denote the agent's utility at time 0 which is defined as

$$V_0^A(\xi) := \sup_a \mathbb{E}^a \left(U_A \left(\xi - \int_0^T k(a_s)\, ds \right) \right) \tag{2.7}$$

recall the definition of k in (2.5). As common in agency problems, we define the concept of incentive-compatible contracts.

Definition 2.1. A contract ξ is said to be *incentive compatible* if V_0^A is finite and if there exists an effort policy $a^*(\xi) \in \mathcal{A}$ that maximizes (2.7), i.e.

$$V_0^A(\xi) = \mathbb{E}^{a^*(\xi)} \left(U_A \left(\xi - \int_0^T k(a_s^*(\xi))\, ds \right) \right).$$

It is critical to understand what incentive-compatible contracts are, as these are the ones for which the principal can enforce desirable efforts. As common in the literature, we will focus on a class Ξ of contracts ξ that are incentive-compatible (IC). Before defining rigorously the class of IC contracts Ξ we will focus on, we clarify the principal's problem. By offering an incentive-compatible contract $\xi \in \Xi$, the principal will be able to anticipate the optimal effort level $a^*(\xi)$. Hence, she will propose an incentive-compatible contract that maximizes the expected value of her CARA preference. Then, her aim is to solve

$$V_0^P := \sup_{\xi \in \Xi} \mathbb{E}^{a^*(\xi)} \left[U_P \left(X_T - \xi \right) \right], \tag{2.8}$$

under the participation constraint $\mathbb{E}^{a^*(\xi)} \left(U_A \left(\xi - \int_0^T k(a_s^*(\xi))\, ds \right) \right) \geq R_0$.
The first result of this paper is given by the following theorem which shows that the problem (2.8) admits an optimal contract which is linear in end-of-period outcomes. The result of the Holmstrom-Milgrom model thus extends to all Gaussian Volterra processes, even though these may exhibit very different statistical properties. Following Schättler and Sung (1993), we introduce the class of contracts we will focus on. Let us define

$$f^*(z) := \frac{\gamma_A}{2}|z|^2 + \inf_{a \in \mathbb{R}} \{k(a) - az\} = \frac{\kappa\gamma_A - 1}{2\kappa} z^2, \tag{2.9}$$

and consider the following class Ξ of Incentive Compatible contracts, see Proposition 3.2 below,

$$\Xi = \{\xi = Y_T^{(y,\beta)} \in \mathcal{F}_T^X, \text{ where } y \geq y_0, \beta = (\beta_t)_{t \leq T} \in \mathcal{A} \text{ and } Y_T^{(y,\beta)} = y + \int_0^T f^*(\beta_s) ds + \int_0^T \beta_s dB_s\}.$$

We have:

Theorem 2.1. *The optimal contract ξ^* that maximizes the principal problem (2.8) is linear in end-of-period profits X_T and is given by*

$$\xi^* = y_0 - \frac{\gamma_P + 1/\kappa}{\gamma_A + \gamma_P + 1/\kappa} g_0(T) + \frac{\kappa\gamma_A - 1}{2\kappa} \int_0^T (\beta_s^*)^2 \, ds + \frac{\gamma_P + 1/\kappa}{\gamma_A + \gamma_P + 1/\kappa} X_T, \quad (2.10)$$

and the optimal level of recommended effort a^* that maximizes the agent's problem (2.7) is deterministic and given by $a^* = \frac{\beta^*}{\kappa}$ with

$$\beta_t^* = \frac{\gamma_P + 1/\kappa}{\gamma_A + \gamma_P + 1/\kappa} K(T, t), \quad t \leq T. \quad (2.11)$$

Proof. See Section 4. □

Similarly to HM, the optimal compensation is made up of a deterministic base salary $y_0 - \frac{\gamma_P+1/\kappa}{\gamma_A+\gamma_P+1/\kappa} g_0(T) + \frac{\kappa\gamma_A+1}{2\kappa} \int_0^T (\beta_s^*)^2 \, ds$ and a random compensation to foster incentives $\frac{\gamma_P+1/\kappa}{\gamma_A+\gamma_P+1/\kappa} X_T$. One of the striking results is, when agents have CARA preferences, the incentive part of the optimal contract, through the performance-based bonus coefficient $\frac{\gamma_P+1/\kappa}{\gamma_A+\gamma_P+1/\kappa}$, is common to all one-dimensional Volterra Gaussian models and thus independent of the output dynamics, even though they have very different statistical properties. Only the base salary is industry-specific depending on the output dynamic through the Volterra kernel K. The optimal effort level is deterministic and firm-specific and can, depending on the choice of the Volterra kernel, exhibit interesting behaviors. For instance, for the mean-reverting dynamics, *i.e.* $K(t, s) = e^{-\lambda(t-s)} \mathbb{1}_{s<t}$, the optimal effort is increasing if the mean-reverting intensity λ is positive. The closer one gets to contract maturity, the more work the agent has to do. The intuition is that the optimal effort should compensate for the natural tendency of the process to revert to its long-term average. The closer the contract is to maturity, the greater the effort should be to allow X to deviate from its long-term average and thus allow the principal to benefit from a greater profit. When the mean-reverting intensity is negative, the effort must be greater at the beginning of the contract in order to give the necessary impetus to the process to diverge towards large positive values. Once this momentum is established, it is less effective to ask the agent to work.

The following two sections are dedicated to proving Theorem 2.1. Section 3 solves the agent problem, while section 4 solves the principal problem. An extension of Theorem 2.1 to the multi-dimensional set-up is considered in Section 5.

3 The one-dimensional agent problem

This section aims at completely solving the problem of the Agent in (2.7). The ideas developed here are not new, they rely on the martingale approach to stochastic control already used in Schättler and Sung (1993) which we adapt to develop the first-order approach to

principal-agent problems in a continuous-time Gaussian setting with exponential utilities. We will show that the class Ξ of contracts are incentive compatible contracts and design the optimal response of the agent for a given contract in Ξ. Our construction relies on the following martingale optimality principle that brings a clear intuition of the stochastic maximum principle used in the context of dynamic contracting by Williams (2011).

3.1 The Martingale optimality principle

The Martingale optimality principle must be seen as a sufficient condition for a contract to be incentive-compatible. The following lemma, which is due to Hu et al. (2005) and proved in Appendix 7.1 for completeness, states this principle.

Lemma 3.1. Given a contract ξ, suppose the existence of a family of stochastic processes $R^a(\xi) := (R^a_t)_{t\leq T}$ indexed by $a \in \mathcal{A}$ such that the following four assertions hold

i) $R^a_T = U_A\left(\xi - \int_0^T k(a_s)ds\right), \quad \forall a \in \mathcal{A},$

ii) $R^a_\cdot$ is a $((\mathcal{F}_t)_{t\in[0,T]}, \mathbb{P}^a)$-supermartingale for every a in $\mathcal{A}$,

iii) R^a_0 is independent of a,

iv) there exists a^* in $\mathcal{A}$, such that R^{a^*} is a $(\mathcal{F}_t)_{t\in[0,T]}$-martingale.

Then, ξ is incentive compatible for the Agent problem (2.7) and a^* is the agent best reply.

In the dynamic agency literature, the process $(R^{a^*}_t)_{t\leq T}$ describes the Agent's expected utility given the contract ξ. A contract ξ is thus incentive compatible if we are able to build such a family $R^a(\xi)$. This will be done in the next section.

3.2 Enlarging the class of Incentive Compatible Contracts

In accordance with the result of Schättler and Sung (1993), we expect that the contracts belonging to Ξ are incentive compatible. It is at this point that a difficulty arises in our setting compared to the Brownian model of Holmstrom and Milgrom (1987) and more generally to the standard literature where the information sets of the two players coincide in the absence of moral hazard. Because the principal has a coarser information (recall that the paths of B are not always observable by the principal), she cannot in general implement the process $(Y^{y,\beta}_t)_{t\leq T}$ given by

$$Y^{y,\beta}_t = y + \int_0^t f^*(\beta_s)ds + \int_0^t \beta_s dB_s, \tag{3.1}$$

for $y \geq y_0$ and $\beta \in \mathcal{A}$, because $Y^{y,\beta}_t$ fails to be $\mathcal{F}^X_t$-measurable. In other words, the contracts in Ξ, that are the most natural to be incentive compatible are a priori inaccessible,

unless we are able to characterize the controls $\beta \in \mathcal{A}$ that induce $Y_T^{(y,\beta)} \in \mathcal{F}_T^X$. Putting aside for a while this problem of information asymmetry between the two players, we consider a larger game where the principal is supposed to have the same information as the agent. We will forget for a while the constraint $\xi \in \mathcal{F}_T^X$ and introduce the enlarged set of contracts

$$\hat{\Xi} = \{\xi = Y_T^{(y,\beta)} \text{ where } y \geq y_0,\, \beta \in \mathcal{A}, \quad Y_T^{(y,\beta)} = y + \int_0^T f^*(\beta_s)ds + \int_0^T \beta_s dB_s\}$$

and naturally extend Definition 2.1 of incentive-compatibility for $\mathcal{F}_T^B$-measurable contracts. We have the following result that we prove for sake of completeness using the Martingale optimality principle.

Proposition 3.2. Let $\hat{\xi} \in \hat{\Xi}$ be of the form

$$\hat{\xi} = y + \int_0^T f^*(\beta_s)ds + \int_0^T \beta_s dB_s,$$

with $y \geq y_0$ and $\beta \in \mathcal{A}$. Then, $\hat{\xi}$ is incentive compatible for the Agent problem in (2.7) and satisfies the participation constraint. Furthermore, the agent best reply is given by the effort $a^*(\hat{\xi}) = \frac{\beta}{\kappa}$ and the utility of the agent at 0 is given by $V_0^A(\hat{\xi}) = -\exp(-\gamma_A y)$.

Proof. Fix $\hat{\xi} \in \hat{\Xi}$. Let $y \geq y_0$ and $\beta \in \mathcal{A}$ such that $\hat{\xi} = y + \int_0^T f^*(\beta_s)ds + \int_0^T \beta_s dB_s$ and define the process $Y^{(y,\beta)}$ by (3.1). For an admissible effort policy $a = (a_t)_{t\leq T} \in \mathcal{A}$, we define R^a as

$$R_t^a := -\exp\left(-\gamma_A\left(Y_t^{(y,\beta)} - \int_0^t k(a_s)ds\right)\right), \quad t \in [0,T].$$

We will show that the family $(R_t^a)_{t\leq T}$ satisfies condition i)–iv) of the Martingale optimality principle of Lemma 3.1. Observe that $Y_T^{(y,\beta)} = \hat{\xi}$ so that Lemma 3.1-i) is satisfied. Also, $R_0^a = -\exp(-\gamma_A y)$ is independent of a as needed in Lemma 3.1-iii). Furthermore, recalling that $B^a = B - \int_0^\cdot a_s\, ds$, we note that

$$\begin{aligned} Y_t^{(y,\beta)} - \int_0^t k(a_s)ds &= y + \int_0^t \left(f^*(\beta_s) - k(a_s)\right) ds + \int_0^t \beta_s dB_s \\ &= y + \int_0^t \left(f^*(\beta_s) - k(a_s) + a_s\beta_s\right) ds + \int_0^t \beta_s dB_s^a. \end{aligned}$$

Using the definition of f^* in (2.9), a completion of the squares in a_s yields the expression

$$f^*(\beta_s) - k(a_s) + a_s\beta_s = -\frac{\kappa}{2}\left(a_s - a_s^*\right)^2 + \frac{\gamma_A}{2}\beta_s^2$$

with $a_s^* := \frac{\beta_s}{\kappa}$ so that combining the above leads to

$$R_t^a = -\exp\left(-\gamma_A y\right)\exp\left(\frac{\gamma_A\kappa}{2}\left(a_s - a_s^*\right)^2\right)\exp\left(-\frac{\gamma_A^2}{2}\int_0^t \beta_s^2 ds - \gamma_A\int_0^t \beta_s dB_s^a\right)$$

It remains to argue that the process $M^a := \exp\left(-\gamma_A \int_0^\cdot \beta_s dB_s^a - \frac{\gamma_A^2}{2}\int_0^\cdot \beta_s^2\, ds\right)$ is a martingale under $\mathbb{P}^a$. Indeed, if this is the case, then, since $-\exp(\frac{\gamma_A\kappa}{2}(a_s - a_s^*)^2 \leq -1$,

$$\mathbb{E}^a[R_T^a] \leq -\exp(-\gamma_A y)\mathbb{E}^a[M_T^a] = -\exp(-\gamma_A y) = R_0^a,$$

which shows that R^a is a $\mathbb{P}^a$-supermartingale for each $a \in \mathcal{A}$, which corresponds to condition Lemma 3.1-ii). Furthermore, for $a = a^*$, we have that R^{a^*} is a $\mathbb{P}^{a^*}$-martingale which gives Lemma 3.1-iv). Obtaining that M^a is a martingale under $\mathbb{P}^a$ is equivalent to proving that

$$M_t = \exp\left(\int_0^t a_s\, dB_s - \frac{1}{2}\int_0^t a_s^2\, ds\right)\exp\left(-\gamma_A\int_0^t \beta_s\, dB_s^a - \frac{\gamma_A^2}{2}\beta_s^2\, ds\right)$$

is a martingale under $\mathbb{P}^0$. But, observe that

$$M_t = \exp\left(\int_0^t (a_s - \gamma_A\beta_s)\, dB_s - \frac{1}{2}\int_0^t (a_s - \gamma_A\beta_s)^2\, ds\right)$$

which is a martingale for $(a_t)_t$ and $(\beta_t)_t$ in $\mathcal{A}$. An application of Lemma 3.1 shows that $\hat{\xi}$ is incentive compatible such that the agent best reply is given by the effort $a_s^*(\hat{\xi}) = \frac{\beta_s}{\kappa}$. Finally, since $y \geq y_0$, the identity

$$V_0^A(\xi) = \mathbb{E}^{a^*}[R^{a^*}] = -\exp(-\gamma_A y) \geq -\exp(-\gamma_A y_0) = R_0,$$

shows that $\hat{\xi}$ satisfies the participation constraint by giving the required utility of the agent at 0, which concludes the proof. □

Notably, when the principal offers a contract parametrized by the pair (y, β), the agent best reply is $\frac{\beta}{\kappa}$ and thus independent of y. This is due to the no wealth effect of CARA preferences. The agent utility is $-\exp(-\gamma_A y)$ and thus independent of β. This is due to the agent's full commitment allowing the principal to choose the best incentive contract that binds the participation constraint. To sum up, restricting our attention to contracts in $\hat{\Xi}$ transforms the puzzling principal's problem to a stochastic Volterra control problem, namely[3]

$$V_{SB} = \sup_{y\geq y_0, \beta\in\mathcal{A}} \mathbb{E}^\beta\left[U_P\left(X_T - Y_T^{y,\beta}\right)\right] = \sup_{\hat{\Xi}} \mathbb{E}\left[U_P\left(X_T - \hat{\xi}\right)\right]. \tag{3.2}$$

where $(Y_t^{y,\beta})_{t\leq T}$ is given by (3.1).

The principal problem (3.2) corresponds to the enlarged stochastic control problem where the principal would have access to the information generated by the Brownian motion. Clearly, the principal value (2.8) satisfies $V_0^P \leq V_{SB}$ because of the inclusion $\Xi \subset \hat{\Xi}$. In the one-dimensional Brownian model, Holmstrom and Milgrom (1987) show that the

[3] To alleviate notations, we will denote hereafter $\mathbb{P}^\beta$, the probability corresponding to the agent's effort choice $a = \frac{\beta}{\kappa}$.

two values coincide because the sets of information $\mathbb{F}^B$ and $\mathbb{F}^X$ are identical ($\Xi = \hat{\Xi}$) and thus there is no need to introduce the enlarged control problem. Our contribution will be to show that the two values always coincide for one-dimensional Gaussian Volterra models, even if $\mathbb{F}^X$ is strictly included in $\mathbb{F}^B$. This is the object of the next section.

4 The one-dimensional principal Gaussian problem

This section is devoted to the explicit resolution of the principal problem (3.2) and to the proof of Theorem 2.1. Contrary to the standard literature, the problem (3.2) is not a Markovian stochastic control problem because the process X_t is not necessarily Markov. More precisely, it corresponds to a stochastic Volterra control problem with the following controlled processes

$$X_t = g_0(t) + \frac{1}{\kappa}\int_0^t K(t,s)\beta_s ds + \int_0^t K(t,s)dB_s^\beta,$$
$$Y_t^{y,\beta} = y + \frac{\kappa\gamma_A + 1}{2\kappa}\int_0^t \beta_s^2 ds + \int_0^t \beta_s dB_s^\beta.$$

We will show that the optimal second-best contract exists and is furthermore $\mathcal{F}_T^X$-measurable. As a consequence, the second-best principal value V_{SB} will coincide with the principal value V_0^P. In other words, our main message is that there is no gain to the principal in acquiring more information than that generated by the observed output process X in one-dimensional Gaussian Volterra models, regardless of the definition of the kernel K. For instance, in the example of discrete observations of Brownian motion, i.e. $K(t,s) = f(t)\mathbb{1}_{s\leq t}$ with f as in (2.3), there is no gain to the principal in increasing the frequency of the discrete observations of the Brownian output.

For $y \geq y_0$ and a control policy $\beta \in \mathcal{A}$, we define $J(y,\beta) = \mathbb{E}^\beta\left[\exp\left(-\gamma_P\left(X_T - Y_T^{y,\beta}\right)\right)\right]$, in order to write the second-best principal problem

$$V_{SB} = \inf_{y\geq y_0} V_{SB}(y), \text{ with } V_{SB}(y) = \inf_{\beta\in\mathcal{A}} J(y,\beta).$$

The rest of the section is dedicated to the proof of Theorem 2.1 that characterizes the optimal control for the principal problem (3.2). The idea of the proof of Theorem 2.1 is to apply again the martingale optimality principle. To do this, we need to introduce a good family of processes indexed by β. Inspired by the agent problem, one possibility would be to consider the following family

$$\exp\left(-\gamma_P\left(X_t - Y_t^{y,\beta}\right)\right).$$

Unfortunately, it may be impossible to apply Itô's formula since the process X may not be a semi-martingale, as in the fractional Gaussian processes case. To get around this problem,

we introduce a new state variable that can be interpreted as a forward price which is a semi-martingale that coincides with X at date T. Let us define the *effort-corrected forward output* process by

$$g_t^\beta(T) = \mathbb{E}^\beta\left[X_T - \frac{1}{\kappa}\int_t^T K(T,u)\beta_u du \mid \mathcal{F}_t\right].$$

Using the output dynamics (2.4) with effort $\beta \in \mathcal{A}$, we have

$$g_t^\beta(T) = g_0(T) + \frac{1}{\kappa}\int_0^t K(T,u)\beta_u du + \int_0^t K(T,u)dB_u^\beta.$$

Then, we observe that the process $(g_t^\beta(T))_{t\leq T}$ is a semi-martingale on $[0,T)$ with dynamics

$$dg_t^\beta(T) = \frac{1}{\kappa}K(T,t)\beta_t dt + K(T,t)dB_t^\beta \tag{4.1}$$

and terminal value $g_T^\beta(T) = X_T$. To apply the martingale optimality principle, we will consider the family of processes

$$M_t^\beta = \exp\left(-\gamma_P\left(g_t^\beta(T) - Y_t^{y,\beta}\right) + \phi_t\right),$$

where ϕ is the deterministic function given by

$$\phi_t = \frac{\gamma_P}{2}\left(\gamma_P^2 - \frac{(\gamma_P + 1/\kappa)^2}{(\gamma_A + \gamma_P + 1/\kappa)}\right)\int_t^T K(T,s)^2 ds.$$

Lemma 4.1, which is proved in Appendix 7.1, provides the dynamics of M^β that plays a key role in the determination of the optimal contract.

Lemma 4.1. *For each $\beta \in \mathcal{A}$, we have*

$$\frac{dM_t^\beta}{M_t^\beta} = \frac{\gamma_P}{2}(\gamma_A + \gamma_P + 1/\kappa)(\beta_t - \beta_t^*)^2 dt + \left(\gamma_P\beta_t - \gamma_P K(T,t)\right)dB_t^\beta, \quad \mathbb{P}^\beta - a.s. \tag{4.2}$$

with β^* given by (2.11).

Proof. See Appendix 7.1. □

We can now complete the proof of Theorem 2.1.

Proof of Theorem 2.1. • *The Principal's problem* is solved by an application of the martingale optimality principle on the process M^β. Fix $\beta \in \mathcal{A}$ and $y \geq y_0$. We show that the family M^β satisfies the four assertions of the martingale optimality principle in Lemma 3.1. We have,

i) For all $\beta \in \mathcal{A}$, it follows from (4.2) that M^β is a $\mathbb{P}^\beta$-sub-martingale and thus

$$M_0^\beta \leq \mathbb{E}^\beta \left[M_T^\beta \right] = \mathbb{E}^\beta \left[\exp\left(-\gamma_P \left(X_T - Y_T^{y,\beta} \right) \right) \right] = J(y, \beta),$$

where we used that $g_T^\beta(T) = X_T$.

ii) Observe that $M_0^\beta = \exp\left(-\gamma_P \left(g_0(T) - y\right) + \phi_0\right) = M_0$ is independent of β.

iii) Finally, for β^* given by (2.11), M^{β^*} is a $\mathbb{P}^{\beta^*}$-martingale by (4.2) and thus, we have

$$J(y, \beta^*) = \mathbb{E}^\beta \left[M_T^{\beta^*} \right] = M_0^{\beta^*} = M_0 \leq J(y, \beta), \quad \beta \in \mathcal{A},$$

which shows that β^* is the optimal control for the second-best principal problem and the principal value is given by $V_{SB}(y) = M_0$.

Optimizing on $y \geq y_0$, yields $V_{SB} = \exp\left(-\gamma_P \left(g_0(T) - y_0\right) + \phi_0\right)$. Furthermore, since β_t^* is proportional to the Volterra Kernel $K(T,t)$, it is straightforward to obtain the linear form of the contract $\xi^* = Y_T^{y_0,\beta^*}$ as in (2.10). In particular, ξ^* is $\mathbb{F}_T^X$-measurable as an affine function of X_T. Therefore, the optimal control for the enlarged principal problem (3.2) induces an optimal contract in Ξ, so that $V_{SB} = V_0^P$.

• *The Agent's problem.* An application of Proposition 3.2 yields that the optimal level of recommended effort a^* that maximizes the agent's problem (2.7) is given by $a^* = \frac{\beta^*}{\kappa}$. □

To sum up, this study shows that the transition from Brownian to Volterra models preserves the optimality of linear in end-of-period profit contracts. Moreover, in one-dimensional models, the principal does not suffer from having a coarser information than the agent about the dynamics of the production process. Aggregating production over time is sufficient for optimal compensation in Volterra Gaussian environments and there is no need to use all the available information - the brownian path in our setting- to design an optimal contract. The next section deals with the robustness of this result in the multi-dimensional set-up.

5 The multi-dimensional model

So far, we have assumed that the shocks are modeled by a one-dimensional Brownian motion. In this section, we present a tractable class of multitask principal-agent problems, such as the one faced by a firm with a manager that supervises several projects. This model amounts to study the principal-agent problem in the case where the shocks are modeled by a standard Brownian motion of dimension d, that we also denote by $(B_t)_{t \leq T}$. As in Holmstrom and Milgrom, the i-th component B_t^i of B_t is interpreted as the outcome of the i-th account of the firm. We model the aggregate output or profit of the firm as follows

$$X_t = g_0(t) + \int_0^t < K(t,s), dB_s >,$$

where $< \cdot, \cdot >$ denotes the canonical inner product in $\mathbb{R}^d$, g_0 is a deterministic function and $K : [0,T]^2 \to \mathbb{R}^d$ is a measurable Volterra Kernel, i.e. $K(t,s) = 0$ for $s \geq t$ such that

$$\sup_{t \leq T} \int_0^T ||K(t,s)||^2 ds < \infty.$$

For instance, the case $K(t,s) = \sigma \mathbb{1}_{s \leq t}$, $\sigma \in \mathbb{R}^d$ corresponds to the Brownian model in Holmstrom and Milgrom (1987, Section 4). We can also consider a mining company that exploits two different types of minerals in two different mines. Each component of X represents the revenue of a mine. This may correspond to the choice $K(t,s) = (e^{-\lambda_1(t-s)}, e^{-\lambda_2(t-s)})$ where each separate outcomes follows a mean-reverting process with two different mean-reverting intensity λ_i, $i = 1, 2$.

Even more importantly than in the one-dimensional case, the filtration generated by the output $\mathbb{F}^X$ is strictly included in the filtration generated by the multi-dimensional Brownian motion $\mathbb{F}^B$ and thus, the principal has always a coarsest information than the agent even when the latter does not exert any hidden effort. This observation is central to Holmstrom and Milgrom's distinction between the optimal contracts that are linear in profits or in accounts and more generally to understand when it is useless to use all the information generated by the Brownian motion. It is useful to recall here that we focus in this paper on contracts that are $\mathcal{F}_T^X$ measurable.

In a similar way to the one-dimensional case, we assume that the agent can exert a continuous vector of effort $(a_t)_{t \leq T} \in \mathbb{R}^d$, a_t^i being the effort made by the manager to improve the account i, that modifies the probability distribution of X as follows:

$$X_t = g_0(t) + \int_0^t < K(t,s), dB_s + a_s \, ds >,$$

Similarly to the one-dimensional case, we say that an agent's action $a = (a_t)_t$ is admissible if $a = (a_t)_t$ is $\mathbb{F}$-progressively measurable and such that there exists $A > 0$ such that

$$\int_0^T ||a_s||^2 \, ds < A, \quad \mathbb{P}^0 - a.s.$$

Still denoting by $\mathcal{A}$ the set of admissible actions, we define for any $a \in \mathcal{A}$ a family of equivalent probability measures $\mathbb{P}_a$ by

$$\frac{d\mathbb{P}_a}{d\mathbb{P}_0} = \exp\left(\int_0^T < a_s, dB_s > - \frac{1}{2} \int_0^T ||a_s||^2 \, ds \right).$$

Under $\mathbb{P}_a$, the process $B^a = B - \int_0^{\cdot} a_s \, ds$ is a $\mathbb{F}-$Brownian motion and the output dynamics is

$$X_t = g_0(t) + \int_0^t < K(t,s), dB_s^a + a_s \, ds > .$$

We also assume that the agent incurs an instantaneous cost $k(a)$ where k is a convex function on $\mathbb{R}^d$ with $k(0_{\mathbb{R}^d}) = 0$. When the kernel is a constant vector, Holmstrom and Milgrom have considered the case where the effort cost function is $k(a) = g(\sum_{i=1}^d a_i)$ with g strictly convex and have showed that the optimal compensation is linear in profit in that case. Nevertheless, this specification does not allow us to determine the optimal effort that the agent must make in each of his tasks. In this section, we will rather consider a quadratic effort cost function

$$k(a) = \frac{1}{2} < a, \Gamma a >,$$

where Γ is a symmetric positive-definite matrix.

When Γ is proportional to the identity matrix, i.e. $\Gamma = \kappa I_d$ for some $\kappa > 0$, we say that the effort cost function is radial, because, in this case, the effort cost function is proportional to the norm of the vector a. A radial cost is to assume that the effort costs are not specific to the different tasks that define the accounts.

In the sequel, we will highlight the interplay between the choice of the matrix Γ and the optimality of linear contracts. In a nutshell, our main results of this section in the multi-dimensional framework can be summarized as follows:

- If Γ is proportional to the identity matrix, then the optimal contract ξ^* is linear in the end-of-period profit X_T. As in the one-dimensional model, the principal does not have to worry about her lack of information to sign an optimal contract.
- For more general matrices Γ, the optimal contract ξ^* is no longer linear in the end-of-period profit X_T. More importantly, ξ^* is not necessarily $\mathcal{F}_T^X$ measurable, meaning that the less-informed principal cannot implement/sign the contract ξ^*. In this situation, we quantify the gap between such contract ξ^* and the best linear contract that can be implemented by the principal. The gap can be interpreted as the value of information.

5.1 The agent's problem

From a methodological viewpoint, there is no hurdle to adapt the techniques developed in Section 3. As a consequence, we will roughly repeat the approach detailed in Section 3 to apply the martingale optimality principle and consider a class of incentive-compatible contracts. To do this, we assume for a while that the principal has access to the Brownian filtration generated by $(B_t)_t$ and can implement the controlled process

$$Y_t^{y,\beta} = y + \int_0^t f^*(\beta_s)dt + \int_0^t < \beta_s, dB_s >, \tag{5.1}$$

with $y \geq y_0$ and $\beta \in \mathcal{A}$, and for $z \in \mathbb{R}^d$,

$$f^*(z) = \frac{\gamma_A}{2}||z||^2 + \inf_a \left(\frac{1}{2} < a, \Gamma a > - < a, z > \right) = \frac{1}{2} < z, \left(\gamma_A I_d - \Gamma^{-1}\right) z >$$

to offer the wage $\xi = Y_T^{y,\beta}$ which is $\mathcal{F}_T^B$ measurable. For a given contract ξ defined by $(y,\beta) \in [y_0,\infty) \times \mathcal{A}$, the agent has to determine his best response $a^*(\xi)$. To apply the martingale optimality principle, we introduce the family of processes indexed by a given by

$$R_t^a = -\exp\left(-\gamma_A\left(Y_t^{y,\beta} - \int_0^t \frac{1}{2} < a_s, \Gamma a_s > ds\right)\right)$$

The first-order condition gives the agent's best effort, $a^*(\xi)_t = \Gamma^{-1}\beta_t$. Furthermore, the agent utility at time 0 is given by $V_0^A(\xi) = R_0^a = -\exp(-\gamma_A y)$. We collect the result in the following proposition, which is the analogue of Proposition 3.2 in the multi-dimensional framework.

Proposition 5.1. Let $\hat{\xi}$ be a contract of the form

$$\hat{\xi} = y + \int_0^T f^*(\beta_s)ds + \int_0^T < \beta_s, dB_s >,$$

with $y \geq y_0$ and $\beta \in \mathcal{A}$. Then, $\hat{\xi}$ is incentive compatible for the Agent problem (2.7) and satisfies the participation constraint. In particular, the agent best reply is given by the effort $a^*(\hat{\xi}) = \Gamma^{-1}\beta$ and the agent utility at time 0 is given by $V_0^A(\hat{\xi}) = -\exp(-\gamma_A y)$.

We can again re-write the Principal's problem as Volterra stochastic optimal control problem on the process

$$Y_T^{y,\beta} = y + \frac{1}{2}\int_0^T < \beta_t, \left(\gamma_A Id + \Gamma^{-1}\right)\beta_t > dt + \int_0^T < \beta_t, dB_t^\beta >,$$

where the stochastic process $B^\beta = (B_t^\beta)_t$ is a d-dimensional Brownian motion under the probability measure indexed by $a_t^* = \Gamma^{-1}\beta_t$ that we will denote hereafter $\mathbb{P}^*$.
Then, without asymmetry of information, the enlarged principal problem is given by

$$V_{SB} = \sup_{y \geq y_0} V_{SB}(y),$$

with

$$V_{SB}(y) = \sup_{\beta \in \mathcal{A}} \mathbb{E}^*\left[U_P\left(X_T - Y_T^{y,\beta}\right)\right] \tag{5.2}$$

and is an upper bound for the principal problem (2.8) with the constraint $\xi \in \mathcal{F}_T^X$.

5.2 The Enlarged Principal problem

The idea is to mimic the methodology developed in details in the one-dimensional case. For this purpose, we reintroduce the effort-corrected forward output

$$g_t^\beta(T) = \mathbb{E}\left[X_T - \int_t^T < K(T,s), \Gamma^{-1}\beta_s > ds | \mathcal{F}_t\right]$$

and apply the martingale optimality principle to the process

$$M_t^\beta = \exp\left(-\gamma_P\left(g_t^\beta(T) - Y_t^{y,\beta}\right) + \phi_t\right),$$

where ϕ is the deterministic function to determine.

The following theorem gives the optimal contract in a multi-dimensional setting for the enlarged principal problem.

Theorem 5.1. *Let Γ be symmetric positive-definite. The optimal level of effort $\Gamma^{-1}\beta^*$ that maximizes the enlarged principal's problem (5.2) is deterministic and β^* is given by*

$$\beta_t^* = \left(\left(\gamma_A + \gamma_P\right) I_d + \Gamma^{-1}\right)^{-1}\left(\gamma_P I_d + \Gamma^{-1}\right) K(T,t), \quad t \leq T. \tag{5.3}$$

The utility of the principal at time 0 is given by

$$V_{SB} = V_{SB}(y_0)$$

with

$$V_{SB}(y) = -\exp\left(-\gamma_P(g_0(T) - y) + \phi_0\right) \tag{5.4}$$

and

$$\phi_0 = \frac{\gamma_P}{2}\langle K_T, \left(\gamma_P I_d - \left(\gamma_P I_d + \Gamma^{-1}\right)\left(\left(\gamma_A + \gamma_P\right) I_d + \Gamma^{-1}\right)^{-1}\left(\gamma_P I_d + \Gamma^{-1}\right)\right) K_T\rangle_{L^2},$$

where $K_T(s) := K(T,s)$ and $\langle f, g\rangle_{L^2} := \int_0^T < f(s), g(s) > ds$. The optimal contract ξ^* that maximizes the principal problem is given by

$$\xi^* = y_0 + \int_0^T f^*(\beta_s^*)ds + \int_0^T < \beta_s^*, dB_s > . \tag{5.5}$$

Proof. See Appendix 7.1. □

We now make two important observations.

Remark 5.2.

- We note that in general, the optimal contract (5.5) is not linear in X_T, indeed the term $\int_0^T < \beta_s^*, dB_s\rangle$ with β^* given by (5.3) cannot be expressed in terms of the integral $\int_0^T < K(T,s), dB_s >$.
- More importantly, ξ^* is measurable with respect to $\mathcal{F}_T^B$ but not necessarily with respect to the smaller filtration $\mathcal{F}_T^X$, which means that such contract cannot be implemented by the less-informed principal.

The following corollary shows that if the cost is radial then the optimal contract is linear in X_T and it can therefore be implemented by the principal. In Section 5.3 below, we study a class of optimal linear implementable contracts for the principal in case the cost is not radial.

Corollary 5.2. *Assume that the effort cost function is radial, i.e.* $\Gamma = \kappa Id$ *for some* $\kappa > 0$. *Then, the optimal level of effort that maximizes the enlarged principal's problem* (5.2) *is deterministic and given by*

$$\beta_t^* = \frac{\gamma_P + 1/\kappa}{\gamma_A + \gamma_P + 1/\kappa} K(T,t).$$

In particular, the optimal contract ξ^* *is linear in profits and given by*

$$\xi^* = y_0 - \frac{\gamma_P + 1/\kappa}{\gamma_A + \gamma_P + 1/\kappa} g_0(T) + \frac{\gamma_A + 1/\kappa}{2} \int_0^T < \beta_s^*, \beta_s^* > ds + \frac{\gamma_P + 1/\kappa}{\gamma_A + \gamma_P + 1/\kappa} X_T.$$

Furthermore, $V_{SB} = V_0^P$.

Proof. The expression for β^* and ξ^* follow directly from Theorem 5.1. In particular, ξ^* is $\mathcal{F}_T^X$-measurable as an affine function of X_T. Therefore, the optimal control for the enlarged principal problem (5.2) induces a contract which is implementable by the principal, so that $V_{SB} = V_0^P$ □

The message of Corollary 5.2 is very simple. When an agent has to allocate his time on several tasks and when his effort cost function is not specific to tasks and thus measured by the norm of the vector a, it is not necessary for the principal to scrutinize the revenues of each activity. It is sufficient to sign a linear contract in the final value of the aggregate profits to give the optimal incentives without regard to the optimal level of information. When the effort cost function is radial, the principal need not observe the paths of individual accounts to offer an optimal compensation that is linear in profits. With this specification, we can disentangle the optimal efforts to allocate to the different tasks because the ith component to the optimal effort is proportional to the ith component of the kernel. In order to illustrate how an agent should optimally allocate his time to the different tasks he has to perform, let us consider the following toy example. A salesperson must visit two clients in two different geographical areas. We assume that the first geographical area generates Brownian outcomes and the second area generates mean-reverting outcomes. The firm's aggregate output is given by

$$X_t = B_t^1 + \int_0^t e^{-\lambda(t-s)}\, dB_s^2, \text{ with } \lambda > 0.$$

While the share of the output that goes to the agent is independent of the parameter λ, the saleperson must differentiate his customers visit. The first customer's visit must be on a constant basis and the second customer's visit must be accelerated as the maturity of the contract approaches.

5.3 The subclass of linear contracts and the value of information

In this section, we no longer assume that the cost is radial, we study the subclass of linear contracts, and we quantify the incurred loss on the utility of the principal. Beyond their simplicity, the main advantage of linear contracts is that they are $\mathcal{F}_T^X$ measurable and can therefore be implemented by the less-informed principal. We will define the value of information as the premium the principal would have to pay to access the agent's information and implement the optimal contract. We will consider contracts as in (5.1) with $y \geq y_0$ but only for controls β in the form

$$\beta_t = bK(T,t), \quad \text{for some } b \in \mathbb{R}.$$

Note that in this case, Proposition 5.1 still apply to ensure that such contracts are Incentive Compatible and the agent best reply is still $\Gamma^{-1}\beta$. Furthermore, for any $b \in \mathbb{R}$ the contract $Y_T^{y,b}$ is by construction linear in X_T and given by

$$Y_T^{y,b} = y + \int_0^T f^*(bK(T,s))ds + b(X_T - g_0(T)),$$

so that b is the share of the output that goes to the agent. In this case, the principal will optimize on (y,b) to find the optimal linear contract

$$V_{lin} = \sup_{y \geq y_0} V_{lin}(y),$$

with

$$V_{lin}(y) = \sup_{b \in \mathbb{R}} \mathbb{E}^* \left[U_P \left(X_T - Y_T^{y,b} \right) \right] \tag{5.6}$$

A direct computation and optimization of the expectation leads to the following result for the optimal linear contract.

Theorem 5.3. *Let Γ be symmetric positive-definite. The optimal level of effort that maximizes the linear principal's problem (5.6) is given by $\Gamma^{-1}\beta^*$ with*

$$\beta_t^* = b^* K(T,t), \quad t \leq T, \tag{5.7}$$
$$b^* = \frac{\langle K_T, \left(\gamma_P I_d + \Gamma^{-1}\right) K_T \rangle_{L^2}}{\langle K_T, \left(\left(\gamma_A + \gamma_P\right) I_d + \Gamma^{-1}\right) K_T \rangle_{L^2}}.$$

The utility of the principal at time 0 is given by

$$V_{lin} = V_{lin}(y_0)$$

with

$$V_{lin}(y) = -\exp\left(-\gamma_P(g_0(T) - y) + \chi_0\right) \tag{5.8}$$

and

$$\chi_0 = \frac{\gamma_P}{2}\langle K_T, \left(\gamma_P I_d - b^* \left(\gamma_P I_d + \Gamma^{-1}\right)\right) K_T\rangle_{L^2}.$$

The optimal linear contract ξ^* that maximizes the linear principal's problem is given by

$$\xi^* = y_0 - b^* g_0(T) + \int_0^T f^*(\beta_s^*)ds + b^* X_T.$$

Proof. Fix $y \geq y_0$, $b \in \mathbb{R}$ and $\beta_t = bK(T,t)$. Then the random variable $X_T - Y_T^{y,b}$ reads

$$\begin{aligned} X_T - Y_T^{y,b} = g_0(T) - y + &< K_T, \left(b\Gamma^{-1} - \frac{b^2}{2}\left(\gamma_A I_d + \Gamma^{-1}\right)\right) K_T >_{L^2} \\ &+ (1-b)\int_0^T < K(T,s), dB_s^\beta > \end{aligned}$$

and is therefore Gaussian under $\mathbb{P}^\beta$. So that a direct computation of the Laplace transform of a Gaussian random variable yields

$$\mathbb{E}^\beta\left[U_P\left(X_T - Y_T^{y,b}\right)\right] = -\exp\left(-\gamma_P(g_0(T) - y) + F(b)\right)$$

with

$$F(b) = < K_T, \left(\frac{\gamma_P^2}{2} I_d - b\left(\gamma_P^2 I_d + \gamma_P \Gamma^{-1}\right) + \gamma_P \frac{b^2}{2}\left((\gamma_A + \gamma_P) I_d + \Gamma^{-1}\right)\right) K_T >_{L^2}.$$

A direct maximization of F on b yields that the optimum is achieved for b^* given by (5.7) and $F(b^*) = \chi_0$. Maximizing on $y \geq y_0$ then gives $V_{lin} = V_{lin}(y_0)$. □

Obviously, when the principal restricts to linear contracts, her utility at 0 satisfies $V_{lin}(y) \leq V_{SB}(y)$. It follows from (5.4) and (5.8) that $\chi_0 \geq \phi_0$. More precisely, one has

$$V_{lin}(y_0) = V_{SB}\left(y_0 + \frac{\chi_0 - \phi_0}{\gamma_P}\right) = V_{SB}(y_0)\exp\left(\chi_0 - \phi_0\right). \tag{5.9}$$

The term $\exp\left(\phi_0 - \chi_0\right) = \frac{V_{SB}(y_0)}{V_{lin}(y_0)}$ lies in $[0,1]$ and can be interpreted as the value of information in the following way: since in general, the principal cannot implement the optimal contract in the enlarged filtration, recall Remark 5.2, she has to restrict to sub-optimal, more simple but implementable contracts. The price to pay when she restricts to linear contracts, is the decrease of her utility by the factor $\exp\left(\phi_0 - \chi_0\right)$, which would correspond to the price to pay to access the optimal contract. Note that the agent utility at time 0 is unchanged compared to the previous section, and is still equal to $\exp(\gamma_A y_0)$ by Proposition 5.1 when the principal proposes the contract (y_0, b^*).

Remark 5.4. *We note that contrary to the one dimensional setting, the coefficient b^* in (5.7) depends in general on the kernel K. For the case of radial costs, i.e. $\Gamma = \kappa I_d$ for some $\kappa > 0$, one recovers from (5.7) that $b^* = (\gamma_P + 1/\kappa)/(\gamma_A + \gamma_P + 1/\kappa)$ which is independent of K. Note also that in this context, $\chi_0 = \phi_0$, so that the value of information vanishes in this case, that is linear contracts are optimal for the Principal's problem, recall Corollary 5.2.*

Having characterized both the fully optimal contract and the optimal linear end-of-period contract, it remains to compare the performances of the two types of contracts. We implement this comparison by studying the sensitivity of the nonnegative difference $\chi_0 - \phi_0$ with respect to the input kernel K and the cost matrix Γ. The smaller the quantity $(\chi_0 - \phi_0)$, the more efficient the implementation of a linear contract, recall (5.9). The next proposition provides an upper bound for the value of information in terms of two quantities: the condition number[4] of the matrix Γ, denoted $Cond(\Gamma)$ and the L^2-norm of the kernel K. The condition number $Cond(\Gamma)$ measures how sensitive is the effort cost function to changes in efforts.

Proposition 5.3. There exists a positive constant C independent of the dimension d, kernel K and terminal time T such that

$$0 \leq \chi_0 - \phi_0 \leq C(Cond(\Gamma) - 1) \int_0^T ||K(T,t)||^2 \, dt. \tag{5.10}$$

Proof. See Appendix 7.1. □

When the cost is radial, then $Cond(\Gamma) = 1$ so that one recovers $\chi_0 = \phi_0$, meaning that linear contracts are optimal, recall Corollary 5.2. When $Cond(\Gamma)$ is close to one, it means that the agent's best reply in terms of effort, solution to the linear equation $\Gamma a^* = \beta$, is not very sensitive to errors in the principal control β. In that case, it is noticeable that the optimal linear contract is nearly optimal regardless of the Volterra process that drives the output.

For convolution kernels of the form $K(T,t) = \mathbb{1}_{t<T} k(T-t)$, $\int_0^T \|K(T,t)\|^2 dt = \int_0^T \|k(t)\|^2 dt$, so that the upper bound in (5.10) shrinks as the horizon of the contract T decreases, suggesting that linear contract seem more performant for short-term relationships compared to long-term relationships. Furthermore, for the exponential kernel $k(t) = e^{-\lambda t}$ with $\lambda \in \mathbb{R}$, we have $\int_0^T \|K(T,t)\|^2 dt = (1 - e^{-2\lambda T})/2\lambda$, thus the higher is the mean-reverting intensity, the smaller the upper bound. For the fractional kernel $k(t) = \sqrt{2H} t^{H-1/2}$ with $H \in (0,1)$, we have $\int_0^T \|K(T,t)\|^2 dt = T^{2H}$.

We now illustrate numerically the value of information $\exp(\phi_0 - \chi_0)$ using Equation (7.3) for exponential and fractional kernels with $d = 2$ and with a diagonal matrix for the cost efforts $\Gamma = \mathrm{diag}(\lambda_1, \lambda_2)$. First, we look at the case of two exponential kernels $k_i(t) = e^{-\rho_i t}$ with different mean reversions[5] $\rho_i \in \mathbb{R}$, $i = 1, 2$.

[4] The condition number of symmetric positive definite matrix S is the ratio $\frac{\lambda_{max}}{\lambda_{min}}$ where λ_{max} (resp. λ_{min}) is the largest (resp. smallest) eigenvalue of S.

[5] Note the change in notation to avoid confusion with the eigenvalues of Γ.

Figure 1 describes a situation where providing one unit of effort for Task 1 is more costly than for Task 2, $\lambda_1 > \lambda_2$, and when the mean-reverting parameters ρ_i vary. The figure shows that linear contracts are more performant for negative and smaller mean reversions, which is the case that is usually of interest in practice. When the intensity of the most expensive task is fixed and positive, the variation of the intensity parameter of the least expensive task has very little effect on the value of the information. More generally, the value of information increases when ρ_2 increases. The linear contracts are very efficient (more than 90%) when the mean-reverting parameter of the most expensive task is one.

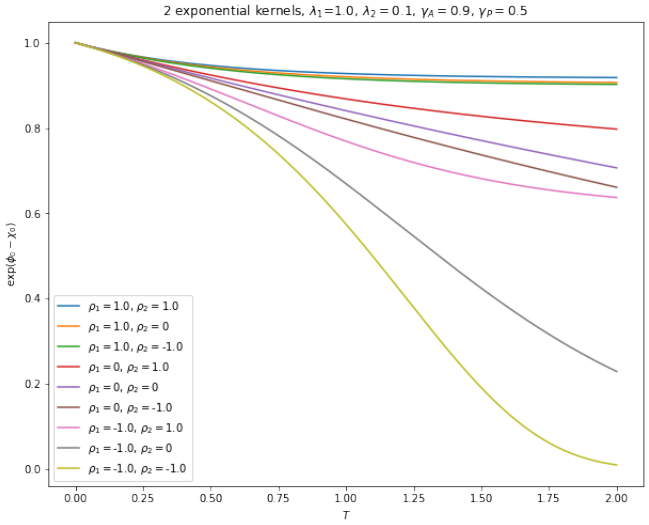


Figure 1: Impact of the mean reversion parameters ρ_1 and ρ_2 on the value of information with respect to the terminal time T.

For our second example, we consider two fractional kernels $k_i(t) = \sqrt{2H_i}t^{H_i-1/2}$ with $H_i \in (0,1)$, we see on Figure 2 that for short maturities linear contracts are more performant for Hurst indices larger than $1/2$ (long-memory processes), while for larger maturities they are more performant for values of $H < 1/2$, (short memory processes). The inflexion point at $T = 1$ is explained by the behavior of the variance of the fractional Brownian motion that reads $\int_0^T \|K(T,t)\|^2 dt = T^{2H}$, see Figure 3.

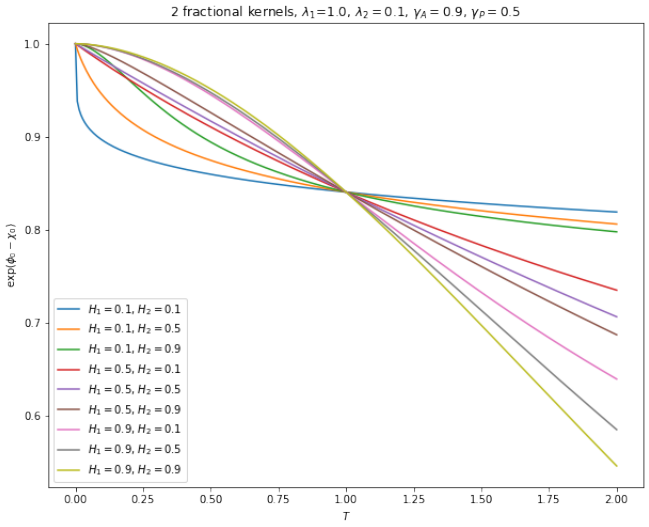


Figure 2: Impact of the Hurst indices H_1 and H_2 on the value of information with respect to the terminal time T.

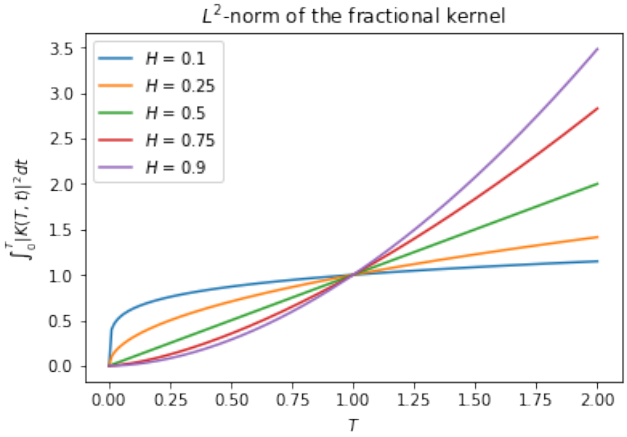


Figure 3: Impact of the Hurst index H on L^2-norm of the fractional kernel $t \mapsto \sqrt{2H}t^{H-1/2}$.

6 Conclusion

The principal-agent framework generally leads to complex optimal contracts that do not align with real-world practices. Also, the proposal of a theoretical framework justifying the signing of simple (linear) optimal contracts deserves attention. In this paper, we have shown that the remarkable results of the Holmstrom and Milgrom model extend surprisingly to a large class of Gaussian models that exhibit memory: the Volterra processes. In particular,

we prove that optimal contracts are linear in one-dimensional models and that the principal has no incentives to expand its information set. In multi-dimensional models, this is no longer generally the case except when the effort cost function is radial. Nevertheless, we are able measure the utility gap when the principal proposes a linear contract in the case of a general effort cost function. Thus, we can examine the key features that make the performance of linear contracts very close to optimality.

References

Abi Jaber, E., Larsson, M., and Pulido, S. (2019). Affine Volterra processes. *The Annals of Applied Probability*, 29(5):3155–3200.

Beran, J. (1994). Statistics for long-memory processes, volume 61 of. *Monographs on Statistics and Applied Probability*, pages 38–72.

Biais, B., Mariotti, T., Plantin, G., and Rochet, J.-C. (2007). Dynamic security design: Convergence to continuous time and asset pricing implications. *The Review of Economic Studies*, 74(2):345–390.

Biais, B., Mariotti, T., Rochet, J.-C., and Villeneuve, S. (2010). Large risks, limited liability, and dynamic moral hazard. *Econometrica*, 78(1):73–118.

Bolton, P. and Dewatripont, M. (2005). *Contract theory*. MIT press.

Carroll, G. (2015). Robustness and linear contracts. *American Economic Review*, 105(2):536–63.

Cisternas, G. (2018). Carrer concerns and the nature of skills. *American Economic Journal: Microeconomics*, 10(2):152–189.

Clementi, G. L. and Hopenhayn, H. A. (2006). A theory of financing constraints and firm dynamics. *The Quarterly Journal of Economics*, 121(1):229–265.

Cvitanić, J., Possamaï, D., and Touzi, N. (2018). Dynamic programming approach to principal–agent problems. *Finance and Stochastics*, 22(1):1–37.

Decreusefond, L. and Ustunel, A. S. (1999). Stochastic analysis of the fractional Brownian motion. *Potential analysis*, 10(2):177–214.

DeMarzo, P. M. and Fishman, M. J. (2007). Optimal long-term financial contracting. *The Review of Financial Studies*, 20(6):2079–2128.

DeMarzo, P. M., Fishman, M. J., He, Z., and Wang, N. (2012). Dynamic agency and the q theory of investment. *The Journal of Finance*, 67(6):2295–2340.

DeMarzo, P. M. and Sannikov, Y. (2006). Optimal security design and dynamic capital structure in a continuous-time agency model. *The Journal of Finance*, 61(6):2681–2724.

Edmans, A. and Gabaix, X. (2011). Tractability in incentive contracting. *The Review of Financial Studies*, 24(9):2865–2894.

Edmans, A., Gabaix, X., Sadzik, T., and Sannikov, Y. (2012). Dynamic ceo compensation. *The Journal of Finance*, 67(5):1603–1647.

Gatheral, J., Jaisson, T., and Rosenbaum, M. (2018). Volatility is rough. *Quantitative Finance*, 18(6):933–949.

Green, E. (1987). Lending and the smoothing of uninsurable income. In Prescott, E. and Wallace, N., editors, *Contractual arrangements for intertemporal trade*, volume 1 of Minnesota studies in macroeconomics, pages 3–25. University of Minnesota Press, Minneapolis.

Gripenberg, G., Londen, S.-O., and Staffans, O. (1990). *Volterra integral and functional equations*, volume 34 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, Cambridge.

Hellwig, M. F. and Schmidt, K. M. (2002). Discrete-time approximations of the holmström-milgrom brownian-motion model of intertemporal incentive provision. *Econometrica*, 70(6):2225–2264.

Holmstrom, B. and Milgrom, P. (1987). Aggregation and linearity in the provision of intertemporal incentives. *Econometrica: Journal of the Econometric Society*, pages 303–328.

Hu, Y., Imkeller, P., and Müller, M. (2005). Utility maximization in incomplete markets. *The Annals of Applied Probability*, 15(3):1691–1712.

Karatzas, I. and Shreve, S. (1991). *Brownian motion and stochastic calculus*, volume 113. Springer.

Kolmogorov, A. N. (1940). Wienersche spiralen und einige andere interessante kurven in hilbertscen raum, cr (doklady). *Acad. Sci. URSS (NS)*, 26:115–118.

Laffont, J.-J. and Martimort, D. (2009). *The theory of incentives*. Princeton university press.

Mandelbrot, B. B. and Van Ness, J. W. (1968). Fractional brownian motions, fractional noises and applications. *SIAM review*, 10(4):422–437.

Rogerson, W. P. (1985). Repeated moral hazard. *Econometrica: Journal of the Econometric Society*, pages 69–76.

Sannikov, Y. (2008). A continuous-time version of the principal-agent problem. *The Review of Economic Studies*, 75(3):957–984.

Schättler, H. and Sung, J. (1993). The first-order approach to the continuous-time principal-agent problem with exponential utility. *Journal of Economic Theory*, 61(2):331–371.

Spear, S. E. and Srivastava, S. (1987). On repeated moral hazard with discounting. *The Review of Economic Studies*, 54(4):599–617.

Sung, J. (1995). Linearity with project selection and controllable diffusion rate in continuous-time principal-agent problems. *The RAND Journal of Economics*, pages 720–743.

Thomas, J. and Worrall, T. (1990). Income fluctuation and asymmetric information: An example of a repeated principal-agent problem. *Journal of Economic Theory*, 51(2):367–390.

Williams, N. (2011). Persistent private information. *Econometrica: Journal of the Econometric Society*, 79(4):1233–1275.

7 Appendix

7.1 Proofs

Proof of Lemma 3.1. Let us consider any admissible effort policy a in $\mathcal{A}$. Conditions i)-iv) immediately imply that

$$\begin{aligned}
&\mathbb{E}^a\left[U_A\left(\xi - \int_0^T k(a_s)ds\right)\right] \\
&\overset{i)}{=} \mathbb{E}^a[R_T^a] \\
&\overset{ii)}{\leq} R_0^a \\
&\overset{iii)}{=} R_0^{a^*} \\
&\overset{iv)}{=} \mathbb{E}^{a^*}[R_T^{a^*}] \\
&\overset{i)}{=} \mathbb{E}^{a^*}\left[U_A\left(\xi - \int_0^T k(a_s^*)ds\right)\right]
\end{aligned}$$

which concludes the proof. □

Proof of Lemma 4.1. Denoting by $U_t^\beta = -\gamma_P \left(g_t^\beta(T) - Y_t^\beta \right) + \phi_t$, an application of Itô's formula yields

$$dM_t^\beta = M_t^\beta \left(dU_t^\beta + \frac{1}{2} d\langle U^\beta \rangle_t \right).$$

Using (4.1), we obtain that

$$dU_t^\beta = \left(\dot{\phi}_t - \gamma_P K(T,t)\kappa\beta_t + \gamma_P \frac{\gamma_A + 1/\kappa}{2} \beta_t^2 \right) dt + \left(\gamma_P \beta_t - \gamma_P K(T,t) \right) dB_t^\beta, \quad \mathbb{P}^\beta - a.s.$$

so that

$$\begin{aligned} \frac{dM_t^\beta}{M_t^\beta} = & \left(\dot{\phi}_t + \frac{1}{2}\gamma_P^2 K(T,t)^2 + \frac{\gamma_P}{2}(\gamma_A + \gamma_P + 1/\kappa)\beta_t^2 - \gamma_P(\gamma_P + 1/\kappa)K(T,t)\beta_t \right) dt \\ & + \left(\gamma_P \beta_t - \gamma_P K(T,t) \right) dB_t^\beta, \quad \mathbb{P}^\beta - a.s. \end{aligned}$$

Completing the squares in β yields

$$\begin{aligned} \frac{\gamma_P}{2}(\gamma_A + \gamma_P + 1/\kappa)\beta_t^2 - \gamma_P(\gamma_P + 1/\kappa)K(T,t)\beta_t = & \frac{\gamma_P}{2}(\gamma_A^2 + \gamma_P + 1/\kappa)\left(\beta_t - \beta_t^* \right)^2 \\ & - \frac{\gamma_P}{2} \frac{(\gamma_P + 1/\kappa)^2}{(\gamma_A + \gamma_P + 1/\kappa)} K(T,t)^2, \end{aligned}$$

with β^* given by (2.11). Combining the above and using that

$$\dot{\phi}_t = \frac{\gamma_P}{2} \left(\frac{(\gamma_P + 1/\kappa)^2}{(\gamma_A + \gamma_P + 1/\kappa)} - \gamma_P^2 \right) K(T,t)^2$$

yields (4.2). □

Proof of Theorem 5.1. The incentive-compatible contract parametrized by (y, β) takes the form

$$dY_t^{y,\beta} = \frac{1}{2} < \beta_t, D\beta_t > dt + < \beta_t, dB_t^\beta >, \quad Y_0^\beta = y,$$

with $D := \gamma_A I_d + \Gamma^{-1}$. Under the probability $\mathbb{P}^\beta$, the output process evolves as

$$X_t^\beta = g_0(t) + \int_0^t < K(t,s), \Gamma^{-1}\beta_s > ds + \int_0^t < K(t,s), dB_s^\beta > .$$

For $s \geq t$, define the effort-corrected forward output by

$$\begin{aligned} g_t^\beta(s) &= \mathbb{E}^\beta \left[X_s^\beta - \int_t^s < K(s,u), \Gamma^{-1}\beta_u > du \mid \mathcal{F}_t \right] \\ &= g_0(s) + \int_0^t < K(s,u), \Gamma^{-1}\beta_u > du + \int_0^t < K(s,u), dB_u^\beta > . \end{aligned}$$

Note that for fixed $s \leq T$, $t \mapsto g_t^\beta(s)$ is a semimartingale on $[0, s)$ with dynamics

$$dg_t^\beta(s) =< K(s,t), \Gamma^{-1}\beta_t > dt + < K(s,t), dB_t^\beta > . \tag{7.1}$$

To mimic the proof of the one-dimensional case, we will apply the martingale optimality principle to the process

$$M_t^\beta = \exp\left(-\gamma_P\left(g_t^\beta(T) - Y_t^\beta\right) + \phi_t\right),$$

with

$$\phi_t = \frac{\gamma_P}{2}\int_t^T < K(T,s), \left(\gamma_P I_d - \left(\gamma_P I_d + \Gamma^{-1}\right)\left(\gamma_P I_d + D\right)^{-1}\left(\gamma_P I_d + \Gamma^{-1}\right)\right) K(T,s) > ds. \tag{7.2}$$

Denoting by

$$U_t^\beta = -\gamma_P\left(g_t^\beta(T) - Y_t^\beta\right) + \phi_t,$$

an application of Itô's formula yields

$$dM_t^\beta = M_t^\beta\left(dU_t^\beta + \frac{1}{2}d\langle U^\beta\rangle_t\right).$$

Using (7.1), we obtain that

$$\begin{aligned} dU_t^\beta = &\left(\dot{\phi}_t - \gamma_P < K(T,t), \Gamma^{-1}\beta_t > + \frac{\gamma_P}{2} < \beta_t, D\beta_t >\right) dt \\ &+ \gamma_P < \beta_t - K(T,t), dB_t^\beta >, \quad \mathbb{P}^\beta - a.s. \end{aligned}$$

so that

$$\begin{aligned} \frac{dM_t^\beta}{M_t^\beta} = &\left[\dot{\phi}_t + \frac{1}{2}\gamma_P^2\|K(T,t)\|^2 + \frac{\gamma_P}{2} < \beta_t, \left(\gamma_P I_d + D\right)\beta_t > \right. \\ &\left. - \gamma_P < K(T,t), (\gamma_P I_d + \Gamma^{-1})\beta_t > \right] dt \\ &+ \gamma_P < \beta_t - K(T,t), dB_t^\beta >, \quad \mathbb{P}^\beta - a.s. \end{aligned}$$

Completing the squares in β yields that

$$\frac{\gamma_P}{2} < \beta_t, \left(\gamma_P I_d + D\right)\beta_t > - \gamma_P < K(T,t), (\gamma_P I_d + \Gamma^{-1})\beta_t >$$

is equal to

$$\begin{aligned} &\frac{\gamma_P}{2} < \left(\beta_t - \beta_t^*\right), \left(\gamma_P I_d + D\right)\left(\beta_t - \beta_t^*\right) > \\ &\quad - \frac{\gamma_P}{2} < K(T,t), \left(\gamma_P I_d + \Gamma^{-1}\right)\left(\gamma_P I_d + D\right)^{-1}\left(\gamma_P I_d + \Gamma^{-1}\right) K(T,t) > \end{aligned}$$

with β^* given by (5.3). Combining the above computations and using (7.2) yield

$$\frac{dM_t^\beta}{M_t^\beta} = \frac{\gamma_P}{2} < (\beta_t - \beta_t^*), (\gamma_P I_d + D)(\beta_t - \beta_t^*) > dt$$
$$+ \gamma_P < (\beta_t - K(T,t)), dB_t^\beta >, \quad \mathbb{P}^\beta - a.s.$$

We then conclude the proof by proceeding analogously to the proof of Theorem 2.1 in Section 4 in the one-dimensional case by applying the Martingale optimality principle on M^β. □

Proof of Proposition 5.3. Let us define the symmetric positive definite matrix

$$A = \gamma_P I_d + \Gamma^{-1}.$$

The matrix A is diagonalizable and thus, there exists an orthogonal matrix P such that

$${}^tPAP = D = diag(\eta_1, \dots, \eta_d)$$

with $\eta_i = \gamma_P + \frac{1}{\lambda_i}$ where $\lambda_1, \dots, \lambda_d$ are the eigenvalues of Γ. We have

$$\chi_0 - \phi_0 = < K_T, A(\gamma_A I_d + A)^{-1} A K > - \frac{< K_T, AK_T >^2}{< K_T, (\gamma_A I_d + A) K_T >}.$$

Setting $K_T(t) = P\hat{K}_T(t)$ for every $t \leq T$ and denoting by $\hat{k}_i(t)$ the components of the vector $\hat{K}_t(t)$, we obtain

$$\chi_0 - \phi_0 = < \hat{K}_T(t), D(\gamma_A I_d + D)^{-1} D \hat{K}_T(t) > - \frac{< \hat{K}_T(t), D\hat{K}_T(t) >^2}{< \hat{K}_T(t), (\gamma_A I_d + D)\hat{K}_T(t) >},$$

or equivalently

$$\chi_0 - \phi_0 = \sum_{i=1}^d \frac{\eta_i^2}{\gamma_A + \eta_i} \int_0^T \hat{k}_i(t)^2 dt - \frac{\left(\sum_{i=1}^d \eta_i \int_0^T \hat{k}_i(t)^2 dt\right)^2}{\sum_{i=1}^d (\gamma_A + \eta_i) \int_0^T \hat{k}_i(t)^2 dt}. \tag{7.3}$$

Then, observing that

$$\sum_{i=1}^d \int_0^T \hat{k}_i(t)^2 dt = \int_0^T \|{}^tPK(T,t)\|^2 dt = \int_0^T \|K(T,t)\|^2 dt,$$

since P is orthogonal, we deduce that

$$\chi_0 - \phi_0 \leq \left(\frac{\eta_{max}^2}{\gamma_A + \eta_{min}} - \frac{\eta_{min}^2}{\gamma_A + \eta_{max}} \right) \int_0^T \|K(T,t)\|^2 dt$$

Rearranging terms and using $a^3 - b^3 = (a-b)(a^2+ab+b^2)$, we finally obtain

$$\chi_0 - \phi_0 \leq C(\eta_{max} - \eta_{min}) \int_0^T \|K(T,t)\|^2 dt,$$

from which we deduce the result by noting that

$$\eta_{max} - \eta_{min} = \frac{1}{\lambda_{max}}(Cond(\Gamma) - 1).$$

□

7.2 Stochastic linear integro-differential convolution equations

In this section, we consider the stochastic linear integro-differential convolution equations of the form

$$dX_t = \left(h(t) + \int_{[0,t]} \mu(ds) X_{t-s} \right) dt + \sigma dB_t, \tag{7.4}$$

with initial condition $X_0 \in \mathbb{R}^d$, where $h : [0,T] \to \mathbb{R}^d$, $\mu : \mathcal{B}([0,T]) \to \mathbb{R}^{d\times d}$ of bounded variation, $X_0 \in \mathbb{R}^d$ and $\sigma \in \mathbb{R}^{d\times d}$. We show that (7.4) admits a unique solution given by the Volterra Gaussian process

$$X_t = g_0(t) + \int_0^t K(t,s) dB_s$$

for some specific choice of input curve $g_0 : [0,T] \to \mathbb{R}^d$ and convolution kernel $K : [0,T]^2 \to \mathbb{R}^{d\times d}$.

Example 7.1. Setting $\mu(dt) = \sum_{k=1}^m a_k \delta_{t_k}$, we recover equations with delay.

For a Lebesgue measurable matrix-valued function f and a matrix-valued measure μ of locally bounded variation we define the convolutions $f * \mu$ and $\mu * f$ by

$$(f*\mu)(t) = \int_{[0,t]} f(t-s)\mu(ds), \quad (\mu * f)(t) = \int_{[0,t]} \mu(ds) f(t-s).$$

We note that $f * \mu = \mu * f$ when $d = 1$.

We need the notion of the differential resolvent of μ. For any bounded measure with locally finite variation μ, the differential resolvent $R : [0,T] \to \mathbb{R}^{d\times d}$ is the unique locally absolutely continuous function R such that

$$R' = \mu * R = R * \mu, \quad R(0) = I_d. \tag{7.5}$$

We refer to (Gripenberg et al., 1990, Theorem 3.1) for the existence and uniqueness statement regarding R. One observes that R is continuous and therefore bounded on $[0,T]$ so that the stochastic convolution $\int_0^\cdot R(\cdot - s)\sigma dB_s$ is well-dedined on $[0,T]$.

Theorem 7.2. *Let $h : [0,T] \to \mathbb{R}^d$, $\mu : \mathcal{B}([0,T]) \to \mathbb{R}^{d\times d}$ of bounded variation and $X_0 \in \mathbb{R}^d$. The stochastic linear integrodifferential equation (7.4) admits a unique (continuous) solution on $[0,T]$ given by*

$$X_t = g_0(t) + \int_0^t R(t-s)\sigma dB_s, \tag{7.6}$$

with

$$g_0(t) = R(t)X_0 + \int_0^t R(t-s)h(s)ds,$$

and R the differential resolvent of μ.

Sketch of proof. • Uniqueness: we show that any solution X to (7.4) is given by (7.6). We first write (7.4) in compact integral form

$$X = X_0 + 1 * h + 1 * (\mu * X) + 1 * \sigma dB.$$

Then, convolving both sides with R, and applying stochastic Fubini's theorem combined with the resolvent equation (7.5), leads to

$$\begin{aligned} R * X &= R * X_0 + R * (1 * h) + R * (1 * \mu * X) + R * (1 * \sigma dB) \\ &= 1 * (RX_0) + 1 * (R * h) + (1 * R * \mu) * X + 1 * (R * \sigma dB) \\ &= 1 * (RX_0) + 1 * (R * h) + R * X - R(0)(1 * X) + 1 * (R * \sigma dB) \end{aligned}$$

Simplifying $R * X$ on both sides, recalling that $R(0) = I_d$, and inspecting the densities lead to

$$X_t = R(t)X_0 + \int_0^t R(t-s)h(s)ds + \int_0^t R(t-s)\sigma dB_s, \quad dt \times d\mathbb{P} - a.e.$$

The conclusion follows from the continuity of the sample paths.

• Existence: we verify that X given by (7.6) solves (7.4). We use the resolvent equation to write $R = I_d + 1 * (\mu * R)$ so that (7.6) reads

$$X = X_0 + 1 * ((\mu * R)X_0 + h + (\mu * R) * h + (\mu * R) * \sigma dB) + 1 * \sigma dB,$$

which shows that X is a continuous semimartingale with the following dynamics

$$\begin{aligned} dX_t &= \left(h(t) + \left(\mu * \left(RX_0 + R * h + R * \sigma dB\right)\right)(t)\right) dt + \sigma dB_t \\ &= \left(h(t) + \left(\mu * X\right)(t)\right) dt + \sigma dB_t \end{aligned}$$

where we made use of stochastic Fubini's theorem. The proof is complete. □