The Aggregation Property 
and its Applications to Realised Higher Moments

Carol Alexander and Johannes Rauch

September 26, 2017

Abstract

We develop a general multivariate aggregation property which encompasses the distinct versions of the property that were introduced by Neuberger [2012] and Bondarenko [2014] independently. This way, we classify new types of model-free realised characteristics for which risk premia may be estimated without bias. We focus on the aggregation property for multivariate martingales and log martingales, and then define realised third and fourth moments which allow long-term higher-moment risk premia to be measured, efficiently and without bias, using high-frequency returns.

Keywords: Aggregation Property, Higher Moments, Risk Premia

*School of Business, Management and Economics, University of Sussex, United Kingdom. Carol Alexander: c.alexander@sussex.ac.uk; Johannes Rauch: j.rauch@sussex.ac.uk.
The aggregation property in finance refers to the invariance of the expected value of a realised characteristic under discretisation of the time interval over which the characteristic is measured. That is, the expected value remains the same irrespective of the partition used to monitor the realised characteristic. For example, if a forward price $F_t$ is a martingale under some measure, then the expected value (under that measure) of the sum of squared changes in the forward price until time $T$ is independent of the sampling frequency. In particular, the expected value is the same – and equal to the expected value of $(F_T - F_0)^2$ – whether the changes are monitored continuously, hourly, daily or weekly over the interval $[0, T]$. The partition of the interval doesn’t even need to be a regular one.

The standard definition of realised variance – an average squared log return – does not satisfy the aggregation property. Therefore the variance risk premium, measured as the difference between the realised variance under the physical measure and the variance swap rate implied from option prices, is biased. By the same token, the theoretical fair value of a conventional variance swap can only be approximated, because the floating leg (realised variance) is computed as the average squared daily log return, whereas the theoretical value assumes the swap is continuously monitored. Consequently, market swap rates can deviate well beyond the no-arbitrage range, especially during crisis periods, which is when trading in volatility products increases. For example, during the financial crisis in 2008, variance swap rates for the Standard & Poor’s 500 Stock Market Index (S&P 500) were frequently 5% or more above the fair value determined by the CBOE Volatility Index (VIX) – see Ait-Sahalia et al. [2015] and Konstantinidi and Skiadopoulos [2016]. The market for variance swaps (and their exchange-traded derivatives) is large, because they are excellent instruments for diversifying investment portfolios and transferring volatility risk. As a result the literature on a variety of discretisation and model-dependent errors in variance swap rates is considerable.

---

1 They were introduced over-the-counter in the 1990’s Demeterfi et al. [1999] and their futures, options, notes, funds and other derivatives are now being actively traded on most large exchanges, with demand stemming from their role as a diversifier, as a hedge, or purely for speculation, as illustrated by Alexander et al. [2015]. Currently, CBOE data show that $3-$6bn notional is traded daily on VIX futures contracts alone and on stock exchanges around the world even small investors can buy and sell over a hundred listed products linked to volatility futures. The most popular of these is Barclay’s VXX note, with a market cap of around $1 trillion.
If one re-defines realised variance so that it satisfies the aggregation property then there exists an exact, model-free fair-value variance swap rate under the minimal assumption of no arbitrage. Also, the same rate applies irrespective of the monitoring frequency of the floating leg. The expectation of the floating leg is path-independent, and even if investors differ in their views about jump risk in an incomplete market they will still agree on the fair-value swap rate. Furthermore, the fair-value swap rate can be expressed in terms of vanilla options written on the underlying, by using the replication theorem of Carr and Madan [2001].

Motivated by the search for more general variance characteristics which satisfy the aggregation property, Neuberger [2012] and Bondarenko [2014] independently provide different definitions for this property. It is not possible to write one in terms of the other, but our paper introduces a general aggregation property which encompasses both definitions, each as a different special case. We characterise a new class of realised pay-offs which satisfy the property, for multivariate martingales and log martingales, including new definitions of higher moments of either price changes or log returns.

The aggregation property (APN) defined by Neuberger [2012] is as follows: given an adapted stochastic process \( x_t \) on a standard filtered probability space and the real-valued function \( g(x) \), the pair \( (g; x) \) satisfies the APN iff

\[
E_r \left[ g(x_T - x_r) \right] = E_r \left[ g(x_s - x_r) \right] + E_r \left[ g(x_T - x_s) \right] \quad \forall \quad 0 \leq r \leq s \leq T , \tag{APN}
\]

where \( E_t \) denotes the expectation under some probability measure conditional on the filtration at time \( t \). Applying the tower law of conditional expectations to (APN) yields

\[
E_0 \left[ \sum_{i=1}^{N} g(\delta x_i) \right] = E_0 \left[ g(x_T - x_0) \right],
\]

for any partition \( \{0 = t_0 < t_1 < \ldots < t_N = T\} \), where \( \delta x_i = x_{t_i} - x_{t_{i-1}} \). That is, if APN holds under a pricing measure, then the realised characteristic on the left has the same market price as the path-independent pay-off on the right.
When \( x = F \) is the forward price of a single tradeable asset the only functions which satisfy \((\text{APN})\) are \( g(\delta F) = \delta F \) and \( g(\delta F) = (\delta F)^2 \), where \( \delta F \) denotes an increment in \( F \). However, Neuberger [2012] focuses on two bivariate cases: the ‘arithmetic’ case, \( x = (F, v) \top \), where \( v_t \) is the conditional variance of \( F_T \); and the ‘geometric’ case, \( x = (y, v^\phi) \top \), where \( y = \ln F \) denotes the log forward price and \( v^\phi_t \) is some ‘generalised’ variance process (defined later in this paper). This allows him to find new second moments which, unlike the sum of squared log returns, satisfy \((\text{APN})\). He also finds one third moment for which \( \text{APN} \) holds and uses this moment to infer the skewness of long-term return distributions from observations on daily log returns.

He concludes by stating “[...] it would also be nice to be able to extend the analysis to higher-order moments. This would not be straightforward.”

Our generalised aggregation property (AP) yields an entire vector space of characteristics which satisfy the AP, including characteristics that allow higher-moment risk premia to be measured without the bias that arises from discretisation and jumps when standard moment definitions are employed.

Bondarenko [2014] introduces an alternative version of the aggregation property (APB) which is based on the levels rather than the increments in a univariate process \( x_t \) and a \( C^1 \) function \( h : \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R} \). He states the property as:

\[
\mathbb{E}_r [h(x_r, x_T)] = \mathbb{E}_r [h(x_r, x_s)] + \mathbb{E}_r [h(x_s, x_T)] \quad \forall \quad 0 \leq r \leq s \leq T. \tag{APB}
\]

He shows that, if \( x = F \) is a martingale (e.g. a forward price), the solutions to APB are given by \( h(x_r, x_s) = a(x_s) - a(x_r) + b(x_r)(x_s - x_r) \) for some real-valued functions \( a \) and \( b \). The \( a \) part corresponds to the trivial solution, since \( h(x_r, x_s) = a(x_s) - a(x_r) \) satisfies APB for any process \( x \). The \( b \) part depends on the martingale assumption and disappears under expectation.

Bondarenko [2014] further examines the special case where \( b = -a' \), see Equation 11 and Corollary 1. This restricted set of functions also appears in Schneider and Trojani [2015], Equation 4, in the context of realised divergence.

---

2See equation (5) p.3430 for the arithmetic case and Proposition 6, p.3435 for the geometric case.

3Kozhan et al. [2013] use the same realised moment to analyse the relationship between the (now unbiased) variance and skewness risk premiums in equity indices, finding that they are closely related.

4Bondarenko [2014] further examines the special case where \( b = -a' \), see Equation 11 and Corollary 1. This restricted set of functions also appears in Schneider and Trojani [2015], Equation 4, in the context of realised divergence.
pair \((g; x)\) satisfies (APN) then \(h(x_r, x_s) = g(x_s - x_r)\) satisfies (APB). However, Bondarenko [2014] leaves the more general case of a vector process for future research, and that is what we present in our paper.

In the following: Section 1 briefly summarises the background literature on conventional variance swaps; Section 2 reviews the results of Neuberger [2012] and Bondarenko [2014] and defines our notation; Section 3 presents our theoretical results on the general aggregation property, characterising an entire vector space of unbiased estimators for the associated risk premia, and then considers the efficiency of these estimators; Section 4 selects some aggregating realised characteristics which correspond to higher moments, focusing on new unbiased and efficient realised third and fourth moments for log returns, and Section 5 concludes. All proofs are in the Appendix.

1 Background on Variance Swaps

A conventional variance swap of maturity \(T\) defines the realised variance (RV) as the average squared daily log return on some underlying over the term of the swap. The calculation of a fair-value swap rate proceeds under the assumptions that the pricing measure is unique\(^5\) and: (a) monitoring of the floating leg happens continuously; (b) the forward price of the underlying follows a pure diffusion process; (c) vanilla options on the underlying with the same maturity as the swap are traded at a continuum of strikes. Then a unique and exact fair-value swap rate – which under assumption (a) becomes the expected quadratic variation (QV) of the log price – is derived from market prices of these options.

However, in the real world none of these assumptions hold. Carr and Wu [2009] discuss the idealised case (a) where the RV becomes the QV of log returns. Then, assuming that the underlying follows a generic jump-diffusion process, they apply the replication theorem of Carr and Madan [2001] to prove that \(\mathbb{E} \left[ \text{QV} \right] = 2 \int_{\mathbb{R}_+} k^{-2}q(k)dk + \iota\), where \(\mathbb{E}\) denotes the

\(^5\)In an arbitrage-free market, as in Harrison and Kreps [1979], expected pay-offs may be computed in a risk-neutral measure. In a complete market the risk-neutral measure for a representative investor corresponds to a unique market implied measure, see Breeden and Litzenberger [1978].
expectation under the pricing measure and $q(k)$ denotes the price of a vanilla out-of-the-money (OTM) option with strike $k$ and maturity $T$. When the underlying price follows a pure diffusion as in (b) the jump error $\iota$ is zero. Regarding assumption (c), in practice the integral must be computed numerically using the vanilla options that are actually traded. Jiang and Tian [2005] address the problems attendant to this assumption and derive upper bounds for the so-called ‘truncation error’. Also based on a finite number of traded strikes, Davis et al. [2014] derive model-free arbitrage bounds for continuously-monitored variance swap rates and claim that market rates are surprisingly close to the lower bound.

A major source of error in the fair-value swap rate stems from assumption (a) because floating legs must be monitored in discrete time. This ‘discrete-monitoring’ error may be written $\varepsilon = \mathbb{E}[RV - QV]$. Then, in the generic jump-diffusion setting of Carr and Wu [2009], the fair-value swap rate for the realised variance may be written $\mathbb{E}[RV] = 2 \int_{-\infty}^{\infty} k^{-2} q(k) dk + \iota + \varepsilon$. There is a large body of research on these pricing errors: Carr and Lee [2009] prove that the discrete monitoring error $\varepsilon$ is related to the third moment of returns; Jarrow et al. [2013] investigate the convergence of the discretely-monitored swap rate to its continuously-monitored counterpart and derive bounds on $\varepsilon$ that get tighter as the monitoring frequency increases; Bernard et al. [2014] generalise these results and provide conditions for signing $\varepsilon$; Hobson and Klimmek [2012] derive model-free bounds for $\varepsilon$; Broadie and Jain [2008] derive fair-value swap rates for discretely-monitored variance swaps under various stochastic volatility diffusion and jump models, claiming that for most realistic contract specifications $\varepsilon$ is smaller than the error due to violation of assumption (b); Bernard and Cui [2014] extend their analysis to include a much wider variety of processes by considering the asymptotic expansion of $\varepsilon$. Finally, Rompolis and Tzavalis [2017] derive bounds for the jump error $\iota$ and demonstrate, via simulations and an empirical study, that price jumps induce a systematic negative bias which is particularly apparent when there are large downward jumps.

However, cutting through this strand of research, both Neuberger [2012] and Bondarenko [2014]...
provide a new class of generalised variance contracts for which exact replication of the floating leg is possible, provided only that the underlying price follows a martingale. The replication strategy consists of a static portfolio of standard options and a dynamic trading strategy in the underlying asset. In these contracts the exposure to variance can also vary over time in response to market conditions, e.g. to increase with the underlying price level. And the key to defining these contracts is the aggregation property.

2 The Aggregation Property

Let \( x_t \in \mathbb{R}^n \) for \( t \in [0, T] \) denote a multivariate adapted process on a standard filtered probability space and consider the derivative process \( u_t = u(t, x_t) \) for a vector-valued function \( u : [0, T] \times \mathbb{R}^n \to \mathbb{R}^n \) as well as a real-valued function \( f : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R} \). This setting allows the unification of the aggregation properties of \cite{Neuberger2012} and \cite{Bondarenko2014} in the single definition as follows:

\[
E_r [f(u_r, u_T)] = E_r [f(u_r, u_s)] + E_r [f(u_s, u_T)] \quad \forall \quad 0 \leq r \leq s \leq T. \tag{AP}
\]

Our aggregation property (AP) is a joint condition on the pair \((f; u)\). Making strong structural assumptions on one gives more flexibility to the other. For example, if \( f(u_r, u_s) = a(u_s) - a(u_r) \) for some function \( a : \mathbb{R}^n \to \mathbb{R} \), all processes \( u \) are a solution; and if \( u \) is constant, then all \( f \) with \( f(u, u) = 0 \) are a solution.

If (AP) holds for \((f; u)\), then by the tower law of expectations

\[
E_0 \left[ \sum_{i=1}^{N} f(u_{i-1}, u_i) \right] = E_0 [f(u_0, u_T)], \tag{E}
\]

for all partitions \( \{0 = t_0 < t_1 < \ldots < t_N = T\} \), where we write \( u_i = u_{t_i} \) for convenience. Following \cite{Neuberger2012}, the interpretation of (E) depends on the measure: if (AP) holds under the physical measure, then \( \sum_{i=1}^{N} f(u_{i-1}, u_i) \) is an unbiased estimator of \( E_0 [f(u_0, u_T)] \) for any partition of \([0, T]\). If (AP) holds under a pricing measure, then the fair price of a
contingent claim that pays $\sum_{i=1}^{N} f(u_{i-1}, u_i)$ is the same as the price of a contingent claim that pays $f(u_0, u_T)$. Under the additional assumption that this contingent claim exists or can be synthesised from other claims, a fair price can be derived from the market.

Our (AP) is more general than (APB) in that we consider a multivariate (not necessarily martingale) derivative process, and more general than (APN) in that our function is defined on the levels at the start and end of an interval in the partition rather than increments $\delta F$ over successive intervals. In fact, it would be possible to write (APB) in terms of (APN), by doubling the size of the state space. That is, setting $x = (F, y)\top$, with $y = \ln F$, we may write

$$g(\delta F, \delta y) = h\left(\frac{\delta F}{e^{\delta y} - 1}, \frac{\delta F e^{\delta y}}{e^{\delta y} - 1}\right), \quad \delta F, \delta y \neq 0.$$ 

However, the induced function $g$ becomes ill-defined as $\delta F, \delta y \to 0$ even when $h$ is a polynomial or some other well-behaved function. Hence, (APN) does not include (APB).

Clearly (APN) is the special case of (AP) where $f(u_r, u_s) = g(x_s - x_r)$. Neuberger [2012] characterises the solutions $(g; x)$ to (APN) when $x$ is bivariate and the second component is a conditional expectation of the first component, whence the process has an implicit dependence structure. First Neuberger [2012] considers the arithmetic case $x = (F, v)\top$, where $v_t = \mathbb{E}_t[(F_T - F_t)^2]$, and finds the solutions

$$g(\delta x) = (c_1, c_2)\top \delta x + c_3 (\delta F)^2 + c_4 \left\{(\delta F)^3 + 3\delta F \delta v\right\},$$

where $\delta x$ denotes an increment in $x$ and $c_i \in \mathbb{R}, \ i \in \{1, \ldots, 4\}$ are arbitrary real coefficients.

The $(c_1, c_2)$ term satisfies aggregation property (APN) trivially, the $c_3$ term occurs provided $F$ is a martingale, since martingales have zero autocorrelation, and the $c_4$ term yields a characteristic which corresponds to a third moment, in the sense that

$$\mathbb{E}_0\left[\sum_{i=1}^{N} \left\{(\delta F_i)^3 + 3\delta F_i \delta v_i\right\}\right] = \mathbb{E}_0\left[(F_T - F_0)^3\right].$$

Then he considers the geometric case $x = (y, v^\phi)\top$, where $y = \ln F$ and $v^\phi_t = \mathbb{E}_t[\phi(y_T - y_t)]$.

\[\text{We thank the associate editor and an anonymous referee for helpful comments in this regard. In fact, Neuberger [2012] doubles the size of the state space, but the conditional variance process } v_t \text{ he includes in } x \text{ cannot be expressed in terms of } F_t, \text{ unlike } y_t.\]
denotes a generalised variance process, i.e. any process such that \( \lim_{\delta y \to 0} \phi (\delta y) / (\delta y)^2 = 1 \).

In this case he shows that the solutions to \((\text{APN})\) are given by

\[
g (\delta x) = (c_5, c_6) \top \delta x + c_7 (e^{\delta y} - 1) + c_8 (\delta v^\phi - 2\delta y)^2 + c_9 (\delta v^\phi + 2\delta y) e^{\delta y},
\]

where \( c_i \in \mathbb{R}, i \in \{5, \ldots , 9\} \), and at least one of \( c_8 \) and \( c_9 \) must be zero. The \((c_5, c_6)\) term satisfies APN trivially and the \( c_7 \) term has zero expectation, provided \( F \) is a martingale. Most interesting are the \( c_8 \) and \( c_9 \) terms: when \( c_8 \neq 0 \) the generalised variance process is called the ‘log variance’ process, denoted \( v_t^\lambda = \mathbb{E}_t [\lambda (y_T - y_t)] \), where \( \lambda (\delta y) = 2 (e^{\delta y} - 1 - \delta y) \).

When \( c_9 \neq 0 \) the generalised variance process is called the ‘entropy variance’ process, denoted \( v_t^\eta = \mathbb{E}_t [\eta (y_T - y_t)] \), where \( \eta (\delta y) = 2 (\delta y e^{\delta y} - e^{\delta y} + 1) \); and when \( c_8 = c_9 = 0 \) then \( v^\phi \) can be any generalised variance process.

Within this geometric set of solutions to \((\text{APN})\), Neuberger [2012] focuses on one particular realised variance, i.e. the log variance (LV), for which \((c_5, c_6, c_7, c_8, c_9) = (-2, 0, 2, 0, 0)\), so that \( g (\delta y) = 2 (e^{\delta y} - 1 - \delta y) = \lambda (\delta y) \).\(^8\) He finds only one higher-order moment, corresponding to \((c_5, c_6, c_7, c_8, c_9) = (6, -3, -12, 0, 3)\), given by

\[
g (\delta x) = \rho (\delta x) + \tau (\delta y),
\]

where \( \rho (\delta x) = 3\delta v^\eta (e^{\delta y} - 1) \) and \( \tau (\delta y) = 6 (\delta y e^{\delta y} - 2e^{\delta y} + \delta y + 2) \).

To find the long-term moment corresponding to this choice of \( g \), set \( \delta x = x_T - x_0 \) and take expectations. Since \( \mathbb{E}_0 [\rho (x_T - x_0)] = 0 \) when \( F \) is a martingale, we have \( \mathbb{E}_0 [g (x_T - x_0)] = \mathbb{E}_0 [\tau (y_T - y_0)] \). Unfortunately, even though \( \lim_{\delta y \to 0} \tau (\delta y) / (\delta y)^3 = 1 \), the implied characteristic does not capture a third moment because it will be dominated by terms in \( \tau \) with order greater than 3 when \( y_T - y_0 \) is sufficiently large. This motivates our search for a new third-moment solution to \((\text{AP})\) for which the long-term moment corresponds to a moment of order exactly 3.\(^9\)

---

\(^8\)Note that \( 2 (\delta y e^{\delta y} - e^{\delta y} + 1) \), i.e. the realised variance corresponding to the entropy variance process, does not satisfy \((\text{APN})\). That is, there is no choice of \((c_5, c_6, c_7, c_8, c_9)\) which yields \( g (\delta y) = \eta (\delta y) \).

\(^9\)The fact that \( \tau (\delta y) \) is \( O (\delta y^3) \) is useful for measurement based on high-frequency data. But \( \rho (\delta x) \) is \( O (\delta v^\eta \delta y) \), so \( g (\delta x) \) is not a pure cubic exposure because it includes an additional price-variance covariance exposure. See Neuberger [2012] for further discussion on this point.
3 Theoretical Results

Our first result characterises the pairs \((f; u)\), with \(f : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}\), which satisfy \((\text{AP})\) as solutions to two second-order partial differential equations, one for the derivative processes \(u = (u_1, \ldots, u_n)^\top\) w.r.t. the underlying processes \(x = (x_1, \ldots, x_n)^\top\) and the other for the real-valued function \(f\) w.r.t. \(u\). For this we need to assume that \(f\) and \(u\) are twice differentiable, so that the following quantities exist: For \(i \in \{1, \ldots, n\}\) let \(\vartheta_t^i = (\vartheta_1^i t, \ldots, \vartheta_n^i t)^\top\) where \(\vartheta_t^i = \vartheta_t^i (t, x_t) \in \mathbb{R}\) denote the time derivatives of \(u\), and let \(\delta_t^i = \delta_t^i (t, x_t) \in \mathbb{R}^n\) and \(\Gamma_t^i = \Gamma_t^i (t, x_t) \in \mathbb{R}^{n \times n}\) denote the first and second partial derivatives of \(u_t\) w.r.t. the components of \(x_t\). For the function \(f\) we denote the Jacobian vector of first partial derivatives w.r.t. the components of the second input vector by \(J = (J_1, \ldots, J_n)^\top \in \mathbb{R}^n\) and write \(H \in \mathbb{R}^{n \times n}\) for the Hessian matrix of second partial derivatives. With these definitions we can now establish necessary conditions for \((f; u)\) to satisfy the \((\text{AP})\), by considering a particular process for \(x_t\) (a multivariate diffusion with a particular drift and covariance) and then deriving conditions for the aggregation property \((\text{AP})\) to hold.

**Theorem 1:** Assume that \(x\) follows a diffusion process with dynamics \(dx_t = \mu_t dt + \Sigma_t dw_t\), where \(\Sigma_t = \Sigma(t, x_t) \in \mathbb{R}^{n \times n}\), \(\mu_t = \mu(t, x_t) \in \mathbb{R}^n\) and \(w_t \in \mathbb{R}^n\) is a standard multivariate Wiener process, for \(t \in [0, T]\). If the first derivatives of \(u\) w.r.t. \(x\), at time \(t\), viz. \(\Delta_t = (\delta_{1t}, \ldots, \delta_{nt})^\top\) form an \(n \times n\) invertible matrix then for every \((f; u)\) such that \(f(u, u) = 0\ \forall \ u \in \mathbb{R}^n\), there exists a function \(a \in C^2\) such that:

\[
H(u_r, u_s) - H^a(u_s) = \sum_{i=1}^n \{J_i(u_r, u_s) - J_i^a(u_s)\} C_{is},
\]

(PDE I)

for any \(0 \leq r \leq s \leq T\), where \(J^a = (J_1^a, \ldots, J_n^a)^\top \in \mathbb{R}^n\) and \(H^a \in \mathbb{R}^{n \times n}\) are the Jacobian and Hessian of \(a\) and \(C_{it} = C_i(u_t) \in \mathbb{R}^{n \times n}\) are symmetric matrices, for \(i = 1, \ldots, n\). Moreover:

\[
2(\vartheta_t + \Delta_t \mu_t) + (\text{tr} \{\Sigma_t^\top M_{1t} \Sigma_t\}, \ldots, \text{tr} \{\Sigma_t^\top M_{nt} \Sigma_t\})^\top = 0,
\]

(PDE II)
where $M_{it} = \Gamma_{it} + \Delta_t^\top C_{it} \Delta_t$, for all $t \in [0, T]$.

Having derived the necessary conditions (PDE I) and (PDE II) we now provide the solutions to (AP) for symmetric matrices $C_{it}$ and derive closed-form expressions for $(f; u)$ in two special cases, viz. Corollary 2, when all $C_{it}$ are constant – so that each process in $u$ follows a log martingale – and Corollary 3 when all $C_{it}$ are zero, in which case $u$ contains only martingales. In each case the definition of the process $u$ is otherwise model-free. Note that the underlying process $x$ does not need to be a diffusion. This assumption was made in Theorem 1 merely to find necessary conditions for the solutions to (AP). Once a general form of solution is found we can verify that it satisfies our AP for any underlying process simply by substitution in (AP). But first we need the following:

**Corollary 1:** According to the multivariate Feynman-Kac formula with boundary condition $u(T, x_T) = \psi(x_T)$ the solution to (PDE II) is given by

$$u(t, x_t) = \mathbb{E}_t \left[ \psi(x_T) + \frac{1}{2} \int_t^T \left( \text{tr} \left\{ \Sigma_{\tau}^\top \Delta_{\tau}^\top C_{1\tau} \Delta_{\tau} \Sigma_{\tau} \right\}, \ldots, \text{tr} \left\{ \Sigma_{\tau}^\top \Delta_{\tau}^\top C_{n\tau} \Delta_{\tau} \Sigma_{\tau} \right\} \right)^\top d\tau \right].$$

Hence, the derivative process $u_t$ follows a martingale if and only if $C_{it} = 0$.

**Theorem 2:** Suppose that $C_{it} = QD_{it}Q^\top$, where $Q$ is orthogonal and $D_{it}$ are diagonal matrices, and that $\sum_{i=1}^n C_{it} = (C_{1t}1, \ldots, C_{nt}1)^\top$. Then the solution to (PDE I) is given by

$$f(u_r, u_s) = a(u_s) - a(u_r) + b(u_r)^\top \left\{ m(u_s) - m(u_r) \right\},$$

where $a : \mathbb{R}^n \to \mathbb{R} \in \mathcal{C}^2$ with $a(0) = 0$, $b : \mathbb{R}^n \to \mathbb{R}^n \in \mathcal{C}^1$, and

$$m(u_t) = \int_0^{u_t} \exp \left\{ \sum_{i=1}^n \int_0^{u_i} C_{i} (\tilde{u}) \, d\tilde{u}_i \right\} d\tilde{u},$$

is a multivariate martingale.
Again the $a$ term satisfies the (AP) trivially, and the $b$ term has zero expectation. If we do not impose $a \in C^2$ and $b \in C^1$ there may be other functions $f$ which satisfy the AP.

The following Corollary characterises all solutions to (PDE I) and (PDE II) in the case that $u$ follows a multivariate log martingale.

**Corollary 2:** Suppose that $C_{it} = C_i$ are constant and that $\sum_{i=1}^{n} C_i$ is invertible. Then

$$m(u_t) = \left( \sum_{i=1}^{n} C_i \right)^{-1} \left( \exp \left( \sum_{i=1}^{n} C_i u_{it} \right) - 1 \right) 1,$$

and

$$u(t, x_t) = \left( \sum_{i=1}^{n} C_i \right)^{-1} \ln \left( \mathbb{E}_t \left[ \exp \left( \sum_{i=1}^{n} C_i \psi_i(x_T) \right) \right] \right) 1,$$

the log of a martingale satisfying the boundary condition $u_T = \psi(x_T)$.

The martingale case in the next Corollary corresponds to the limit as $C_i \to 0$, $i = 1, \ldots, n$ of the log martingale case in Corollary 2. It is straightforward to verify that the martingale condition (M) below characterises all pairs $(f; u)$ which satisfy (AP) when $u$ is a martingale, for any $C^2$ function $a$ and any set of $C^1$ functions in $b$.

**Corollary 3:** Suppose that $C_i = 0$ for all $i \in \{1, \ldots, n\}$. Then the solution to (PDE I) is

$$f(u_r, u_s) = a(u_s) - a(u_r) + b(u_r)^T (u_s - u_r),$$

(M)

where we assume w.l.o.g. that $J^a(0) = 0$. Furthermore, the solution to (PDE II) is

$$u(t, x_t) = \mathbb{E}_t[\psi(x_T)].$$

Note that the univariate martingale case corresponds to the solutions described in equation (12) of Bondarenko [2014] to the aggregation property (APB).
The realised characteristic based on \((f; \mathbf{u})\), i.e.
\[
\sum_{i=1}^{N} f(\mathbf{u}_{i-1}, \mathbf{u}_i) = a(\mathbf{u}_T) - a(\mathbf{u}_0) + \sum_{i=1}^{N} b(\mathbf{u}_{i-1})^\top (\mathbf{u}_i - \mathbf{u}_{i-1}),
\]
is an unbiased estimator for the implied characteristic \(E_0[f(\mathbf{u}_0, \mathbf{u}_T)]\) which, by virtue of \(\mathbf{u}\) being a martingale, may also be written \(E_0[a(\mathbf{u}_T)] - a(\mathbf{u}_0)\). Note that \(\mathbf{u}_i = \mathbf{u}_t\). Thus, the implied characteristic is defined by \(a\) alone. In the next section we shall consider some particular choices for \(a\) which correspond to higher moments.

While \(a\) must be fixed at time 0, \(b\) can change dynamically over time. Moreover, \(b\) determines the conditional variance of the estimator along the partition, because
\[
\mathbb{E}_t \left[ \left\{ \sum_{i=1}^{N} f(\mathbf{u}_{i-1}, \mathbf{u}_i) - \mathbb{E}_t \left[ \sum_{i=1}^{N} f(\mathbf{u}_{i-1}, \mathbf{u}_i) \right] \right\}^2 \right] = \mathbb{E}_t \left[ \left\{ a(\mathbf{u}_T) - \mathbb{E}_t[a(\mathbf{u}_T)] + \sum_{t_i > t} b(\mathbf{u}_{i-1})^\top (\mathbf{u}_i - \mathbf{u}_{i-1}) \right\}^2 \right],
\]
for \(t \in \{t_0, \ldots, t_N\}\). Next we propose an optimal choice for \(b\) in the sense that it yields a conditionally efficient estimator, i.e. an estimator with minimum conditional variance:

**Theorem 3:** Given a martingale process \(\mathbf{u}\) and some function \(a \in \mathcal{C}^2\) which specifies an implied characteristic for this process, the conditionally efficient estimator for the implied characteristic \(E_0[f(\mathbf{u}_0, \mathbf{u}_T)]\) has a first-order approximation given by \(b^* (\mathbf{u}) = -J^a(\mathbf{u})\). If the process is a diffusion and monitoring happens continuously, the approximation is exact.

Theorem 3 gives a first-order approximation which becomes more exact as the monitoring frequency of the realised characteristic increases. The approximation error is not as detrimental to accurate measurement of a risk premium as violations of the AP. On a large sample small violations of \([\text{AP}]\) can result in a large cumulative bias, whereas finding efficiency only to a first-order approximation may be acceptable on a large sample with frequent monitoring.
4 Examples of Aggregating Realised Higher Moments

First we show that the solution \(M\) encompasses the realised variances and third moment found by Neuberger [2012]. To this end we first re-write Neuberger’s geometric set of solutions in an unrestricted fashion by setting \(x = (y, v^\lambda, v^\eta)\) and writing

\[
g(\delta x) = (c_5, c_6^\lambda, c_6^\eta)^\top \delta x + c_7 \left( e^{\delta y} - 1 \right) + c_8 \left( \delta v^\lambda - 2 \delta y \right)^2 + c_9 \left( \delta v^\eta + 2 \delta y \right) e^{\delta y}.
\]

Next we define the log contract and entropy contracts, viz. \(Y_t = \mathbb{E}_t[y_T]\) and \(Z_t = \mathbb{E}_t[F_T y_T]\), respectively. Note that \(v^\lambda = 2(y - Y)\) and \(v^\eta = 2(Z - y)\).

**Corollary 4:** Let \(u = (F, Y, Z)^\top\) and set \(a(u) = (c_5 + 2c_6^\lambda - 2c_6^\eta, 4c_8, 2c_9^\lambda)^\top (\ln F, Y^2, Z^2)^\top\) and \(b(u) = (c_7^\lambda F - 2c_9 Z, -2c_6^\lambda, -8c_8 Y, 2c_9^\lambda F)^\top\). Then

\[
f(u_r, u_s) = a(u_s) - a(u_r) + b(u_r)^\top (u_s - u_r) = g(x_s - x_r),
\]

and hence the solution to \((APN)\) can be expressed in terms of the solution to \((AP)\).

For example, setting \((c_5, c_6^\lambda, c_6^\eta, c_7, c_8, c_9) = (-2, 0, 0, 2, 0, 0)\) so that \(a(u) = -2 \ln F\) and \(b(u) = (2 F, 0, 0)^\top\), we have \(f(u_r, u_s) = \lambda(y_s - y_r)\), which corresponds to Neuberger’s log variance (LV). Similarly, setting \((c_5, c_6^\lambda, c_6^\eta, c_7, c_8, c_9) = (6, 0, -3, -12, 0, 3)\) yields \(a(u) = 12 \ln F - 6 \frac{Z}{F}\) and \(b(u) = (\frac{-12}{F^2} - \frac{6Z}{F^2}, 0, 0, 6)^\top\), and then we have \(f(u_r, u_s) = \rho(x_s - x_r) + \tau(y_s - y_r)\), which corresponds to Neuberger’s third moment (NTM). In both examples \(b(u) = b^*(u)\), so both NTM and LV are approximately efficient. However, in the empirical work of Neuberger [2012] and Kozhan et al. [2013] the vector \(b\) is fixed monthly rather than rebalanced at the daily monitoring frequency, which makes the estimator less efficient.

---

10Expressing Bondarenko’s generalised and power-price weighted variance contracts (p.88) in our notation is a straight-forward task, since his set of solutions corresponds to our univariate martingale case.
Next we suppose that \( u \) contains a series of power log contracts \( P_t^{(i)} = \mathbb{E}_t [y_T^i], \ i \geq 0 \) on a single underlying \( F \). According to the replication theorem of [Carr and Madan 2001], for \( i \geq 1 \) this conditional expectation can be expressed in terms of vanilla out-of-the-money (OTM) options as:

\[
P_t^{(i)} = y_t^i + \int_{\mathbb{R}^+} \gamma_i(k) q_t(k) dk,
\]

where \( \gamma_i(k) := i(\ln k)^{i-2} k^{-2} [i - 1 - \ln k] \) and \( q_t(k) \) denotes the time-\( t \) price of a vanilla OTM option with strike \( k \) and maturity \( T \). In particular, for \( i = 1 \), this yields the replication portfolio \( Y_t = y_t - \int_{\mathbb{R}^+} k^{-2} q_t(k) dk \) of the log contract. So, let \( u = (Y, P^{(2)} \ldots, P^{(n)})^\top \) and consider the specification

\[
a(u) = n (-Y)^{n+1} - \sum_{i=2}^{n} \binom{n+1}{i} P^{(i)} (-Y)^{n+1-i}.
\]

Note that \( a(u_T) = y_T^{n+1} \), since \( P_T^{(i)} = y_T^i \) and \( \sum_{i=0}^{n+1} \binom{n+1}{i} (-1)^{n+1-i} = 0 \). Then, according to Corollary 3, and using that \( P^{(0)} = 1 \) as well as \( P^{(1)} = Y \), the implied characteristic equals

\[
\mathbb{E}_0 [f(u_0, u_T)] = \mathbb{E}_0 [a(u_T)] - a(u_0) = \sum_{i=0}^{n+1} \binom{n+1}{i} P^{(i)}_0 (-Y_0)^{n+1-i} = \mathbb{E}_0 [(y_T - Y_0)^{n+1}] \quad (CM)
\]

which is the \( n + 1^{st} \) central moment of the log return distribution.

For \( n = 1 \) we have \( a(u) = Y^2 \) and, using Theorem 3, \( b^* (u) = -2Y \), so that \( f(u_r, u_s) = (Y_s - Y_r)^2 \), the squared change in price of the log contract. This characteristic corresponds to the \( c_8 \) term in Neuberger’s geometric set of solutions and it can be replicated by holding a squared log contract and shorting \( 2Y \) log contracts from time \( r \) to time \( s \). Here, and in the following, \( s - r = 1/250 \) for daily monitoring.

Note that the calculation of \( Y_t \) from option prices is subject to a numerical integration error, whereas the LV may be derived from direct observation of the underlying price alone. For this reason, the LV is preferable to the squared change in price of the log contract for the unbiased (and efficient) estimation of a variance risk premium. However, there is very little difference between them in Figure 1. Here we compare the new RV estimator with
Figure 1: Comparison of our efficient RV, Neuberger’s LV as well as the conventional squared log return (CONV) for $s - r = 1/250$ (daily monitoring) and $\sigma = 20\%$ (implied volatility). The $x$-axis corresponds to $Y_s - Y_r$, the change in price of the log contract. Note that $y_s - y_r = Y_s - Y_r + \sigma^2 \frac{s - r}{2}$.

existing definitions of realised variance, under the assumption that vanilla options with a continuum of strikes can be traded and therefore the log contract can be synthesised, using the replication theorem of Carr and Madan [2001]. Note that this illustration is otherwise model-free, e.g. it allows for stochastic volatility or jumps in the underlying price process.

For $n \geq 2$ we apply Theorem 3 to derive the first-order efficient estimator for the $n + 1^{st}$ central moment of the log return distribution as:

$$b^\star (u) = \left( n (n + 1) (-Y)^n - \sum_{i=2}^{n} \binom{n+1}{i} (n + 1 - i) P^{(i)} (-Y)^{n-i}, \ldots, (n + 1) (-Y) \right)^\top. \quad (b)$$

Now we use this to define a new realised third moment that does not suffer the same problem as NTM. More precisely, we seek an aggregating third moment for which the implied characteristic corresponds exactly to the third central moment of the log return distribution. Setting $n = 2$ in (a) and (b) yields $a(u) = -2Y^3 + 3P^{(2)}Y$ as well as $b^\star (u) = (6Y^2 - 3P^{(2)}, -3Y)^\top$, and therefore

$$f(u_r, u_s) = (Y_s - Y_r)^3 + 3 (v_s^{(2)} - v_r^{(2)}) (Y_s - Y_r), \quad \text{(RTM)}$$
which is similar to the $c_4$ term in Neuberger’s arithmetic set of solutions, but has the log contract rather than the forward price as the underlying. The pay-off can be replicated by holding a cubed log contract as well as $6Y_r^2 - 3P_r^{(2)}$ log contracts and shorting $3Y_r$ squared log contracts from time $r$ to time $s$.

![Figure 2: Comparison of RTM, Neuberger’s NTM and the cubed log return (CONV), i.e. $(y_s - y_r)^3$, for $s - t = 1/250$ (daily monitoring) and $\sigma = 20\%$ (implied volatility). Note that $\nu_0 \approx \nu^{(2)} + \nu^{(3)}/3$, see Equation 14 in Neuberger [2012]. However, the graph for NTM does not change significantly for reasonable (or indeed unreasonable) values of the implied third moment so we set $\nu^{(3)} = 0$ for simplicity. Again, the x-axis corresponds to $Y_s - Y_r = y_s - y_r - \frac{\sigma^2}{2} (s - r)$.]

The RTM arises from two sources: the cubed change in price of the log contract (i.e. the short-term third moment) and the product of the change in the log contract and the change in conditional variance (i.e. leverage). Following the example of Neuberger [2012] (p.3430, last paragraph), if $F$ is a continuously-sampled continuous martingale, then the cubic term goes to zero and the only remaining source of long-term third moment is leverage. Both RTM and NTM require the replication of synthetic contracts and are therefore subject to a measurement error. The advantage of our definition is that, by (CM), the swap rate corresponds exactly to the third central moment of the log return distribution.

In Figure 2 we compare the two realised third moment measures with the cubed log return $(y_s - y_r)^3$, under the assumption that vanilla options with a continuum of strikes can
be traded and therefore that both the log and squared log contracts can be synthesised. As expected, the RTM is closer to the cubed log return (CONV) than NTM. Note that RTM is marginally below the cubed log return when $Y_s - Y_r$ is positive because the second, leverage term in RTM is positive; and above it when $Y_s - Y_r$ is negative. The difference between RTM and NTM increases with the magnitude of $Y_s - Y_r$ and, for large positive or negative changes, NTM is quite far above the conventional third moment.

Finally we set $n = 3$ in (a) and (b) to obtain a fourth-moment characteristic with $a(u) = 3Y^4 - 6P^{(2)}Y^2 + 4P^{(3)}Y$ as well as $b^*(u) = (-12Y^3 + 12P^{(2)}Y - 4P^{(3)}, 6Y^2, -4Y)^\top$, and therefore (again, with $s - r = 1/250$ for daily monitoring):

$$f(u_r, u_s) = (Y_s - Y_r)^4 + 6v_{s}^{(2)} (Y_s - Y_r)^2 + 4 (v_{s}^{(3)} - v_{r}^{(3)}) (Y_s - Y_r). \quad \text{(RFM)}$$

Intuitively, this realised fourth moment (RFM) derives from three terms: the first is the fourth power of the change in price of the log contract (i.e. short-term fourth-moment); the second is a volatility clustering factor and the third is a leverage term. As with the third moment, the first term is zero if $F$ is a continuously-sampled continuous martingale. The volatility clustering factor is always positive and tends to decrease with residual time to maturity (for $s = T$ it is zero). If the implied distribution of the log price is symmetric then the leverage factor is zero, and then volatility clustering is the only source of kurtosis, if monitoring is continuous. The RFM can be replicated by holding a fourth-power log contract as well as $6Y_r^2$ squared log contracts and shorting $12Y_r^3 - 12P_r^{(2)}Y_r + 4P_r^{(3)}$ log contracts and $4Y_r$ cubed log contracts from time $r$ to time $s$.

Figure 3 compares our realised fourth-moment characteristic with the conventional moment, i.e. the fourth power of the log return. By (CM), the former is constructed to: correspond exactly to an implied moment which captures the fourth central moment of log returns; to satisfy the AP and thus be an unbiased estimator for the realised fourth moment, for any monitoring frequency; and to be the most efficient of these unbiased estimators under high-frequency monitoring. We argue that it has better properties than the conventional mo-
Figure 3: Comparison of RFM with the fourth power of log return (CONV), i.e. \((y_s - y_r)^4\), for \(s - r = 1/250\) (daily monitoring) and \(\sigma = 20\%\) (implied volatility). The graph for RFM does not change significantly for reasonable (or indeed unreasonable) values of implied third moment so we set \(v^{(3)} = 0\) for simplicity. Again, the \(x\)-axis corresponds to \(Y_s - Y_r = y_s - y_r - \frac{\sigma^2}{2}(s - r)\).

ment, which does not satisfy the AP and, as a result, does not allow one to infer the values of long-term fourth moments from short-term observations. We note from the figure that there can be a substantial difference between our fourth moment definition and the conventional definition and that this difference increases with \(|Y_s - Y_r|\).

5 Conclusions

Our general property encompasses two aggregation properties that were independently introduced by Neuberger [2012] and Bondarenko [2014] as distinct examples, the former corresponding to a particular bivariate function and the latter being the univariate, martingale case of our general property. Our initial results are not confined to martingales, or even to log martingales, but in these cases we are able to define new, aggregating characteristics which correspond exactly to higher moments of log returns.

To estimate a risk premium one takes the difference between the risk-neutral characteristic that is implied from traded options and the realised characteristic in the physical
measure, which is typically derived from high-frequency historical data. But if the realised characteristic does not satisfy the aggregation property, the risk premium estimator will be biased – just like the standard variance risk premium estimator. Moreover, the aggregation property allows one to derive unbiased estimators for the realised characteristic independent of the monitoring frequency. Unless the estimator satisfies the aggregation property, it is not possible to infer accurate values of long-term premia from short-term observations.

While the unbiased property is independent of the monitoring frequency, one also needs to consider efficiency. Otherwise, the single historical time series that is observed on returns may, by chance, yield a realised moment estimate which is far from its expectation, even though the estimator is unbiased. So within our vector space of unbiased estimators for higher moments we derive the most efficient, i.e. those with minimum conditional variance, where the efficiency of our selected estimators increases with the monitoring frequency.

We hope that this sets an agenda for further research. For instance, when based upon our aggregating third and fourth moment characteristics, an empirical examination of the determinants of high-moment risk premia may draw different conclusions to previous research. In particular, Kozhan et al. [2013] extends the work of Carr and Wu [2009], Egloff et al. [2010] and others on the determinants of the variance risk premium, only to conclude that the third-moment risk premium is very highly correlated with the variance risk premium. However, it may be that our characteristics which, unlike the aggregating third moment found by Neuberger [2012], correspond exactly to the $n^{th}$ central moment of the implied distribution, do indeed yield diversified risk premia. Such a finding would be important for finance practitioners that seek new and profitable forms of tradable risk.
References

Y. Ait-Sahalia, M. Karaman, and L. Mancini. The term structure of variance swaps and risk premia. *Working Paper*, 2015.

C. Alexander, J. Kapraun, and D. Korovilas. Trading and investing in volatility products. *Financial Markets, Institutions & Instruments*, 24(4):313–347, 2015.

G. Bakshi, N. Kapadia, and D. Madan. Stock return characteristics, skew laws and the differential pricing of individual equity options. *Review of Financial Studies*, 16(1):101–143, 2003.

C. Bernard and Z. Cui. Prices and asymptotics for discrete variance swaps. *Applied Mathematical Finance*, 21:140–173, 2014.

C. Bernard, Z. Cui, and D. Mcleish. Convergence of the discrete variance swap in time-homogeneous diffusion models. *Quantitative Finance Letters*, 2(1):1–6, 2014.

O. Bondarenko. Variance trading and market price of variance risk. *Journal of Econometrics*, 180:81–97, 2014.

D. T. Breeden and R. H. Litzenberger. Prices of state-contingent claims implicit in option prices. *Journal of Business*, 51(4):621–651, 1978.

M. Broadie and A. Jain. The effect of jumps and discrete sampling on volatility and variance swaps. *International Journal of Theoretical and Applied Finance*, 11(8):761–979, 2008.

P. Carr and R. Lee. Volatility derivatives. *The Annual Review of Financial Economics*, 1:1–21, 2009.

P. Carr and D. Madan. Optimal positioning in derivative securities. *Quantitative Finance*, 1(1):19–37, 2001.

P. Carr and L. Wu. Variance risk premiums. *Review of Financial Studies*, 22(3):1311–1341, 2009.

M. Davis, J. Obloj, and V. Raval. Arbitrage bounds for weighted variance swap prices. *Mathematical Finance*, 24(4):821–854, 2014.

K. Demeterfi, E. Derman, M. Kamal, and J. Zou. A guide to volatility and variance swaps. *Journal of Derivatives*, 6(4):9–32, 1999.

D. Egloff, M. Leippold, and L. Wu. The term structure of variance swap rates and optimal variance swap investments. *Journal of Financial and Quantitative Analysis*, 45(5):1279–1310, 2010.

J. M. Harrison and D. M. Kreps. Martingales and arbitrage in multiperiod securities markets. *Journal of Economic Theory*, 20:381–408, 1979.

D. Hobson and M. Klimmek. Model independent hedging strategies for variance swaps. *Finance and Stochastics*, 16:611–649, 2012.
R. Horn and C. Johnson. *Matrix Analysis*. Cambridge University Press, 1985.

R. Jarrow, Y. Kchia, M. Larsson, and P. Protter. Discretely sampled variance and volatility swaps versus their continuous approximations. *Finance and Stochastics*, 17:305–324, 2013.

G. Jiang and Y. Tian. The model-free implied volatility and its information content. *Review of Financial Studies*, 18(4):1305–1342, 2005.

E. Konstantinidi and G. Skiadopoulos. How does the market variance risk premium vary over time? Evidence from S&P 500 variance swap investment returns. *Journal of Banking and Finance*, 62:62–75, 2016.

R. Kozhan, A. Neuberger, and P. Schneider. The skew risk premium in the equity index market. *Review of Financial Studies*, 26(9):2174–2203, 2013.

A. Neuberger. Realized skewness. *Review of Financial Studies*, 25(11):3423–3455, 2012.

L. S. Rompolis and E. Tzavalis. Retrieving risk neutral moments and expected quadratic variation from option prices. *Review of Quantitative Finance and Accounting*, 48(4):955–1002, 2017.

P. Schneider and F. Trojani. Divergence and the price of uncertainty. *Swiss Finance Institute Research Paper*, 15(60), 2015.
6 Appendix

Proof of Theorem 1: Decomposing $f(u_r, u_s)$ using Itô integrals yields

$$f(u_r, u_s) = \int_r^s J(u_r, u_t)^\top du_t + \frac{1}{2} \int_r^s \text{tr} \{ H(u_r, u_t) d\langle u \rangle_t \}. \quad (1)$$

Applying (1) to all terms in (AP), and writing $\hat{J}_t = J(u_r, u_t) - J(u_s, u_t)$ as well as $\hat{H}_t = H(u_r, u_t) - H(u_s, u_t)$ yields

$$\mathbb{E}_r \left[ \int_s^T \hat{J}_t^\top du_t + \frac{1}{2} \int_s^T \text{tr} \left( \hat{H}_t d\langle u \rangle_t \right) \right] = 0, \quad (2)$$

Since $dx_t = \mu_t dt + \Sigma_t dw_t$, Itô’s lemma applied to $u_t = u(t, x_t)$ yields

$$du_t = \lambda_t dt + \Delta_t \Sigma_t dw_t, \quad (3)$$

where

$$\lambda_t = \vartheta_t + \Delta_t \mu_t + \frac{1}{2} \left( \text{tr} \left\{ \Sigma_t^\top \Gamma_1 \Sigma_t \right\}, \ldots, \text{tr} \left\{ \Sigma_t^\top \Gamma_n \Sigma_t \right\} \right)^\top. \quad (4)$$

The quadratic variation of $x$ is $d\langle x \rangle_t = \Sigma_t \Sigma_t^\top dt$, and so

$$d\langle u \rangle_t = \Delta_t \Sigma_t \Sigma_t^\top \Delta_t^\top dt. \quad (5)$$

Inserting (3), (4) and (5) in (2), and noting that the stochastic integral w.r.t. $dw$ vanishes under expectation, since $w$ is a martingale, yields

$$\mathbb{E}_r \left[ \int_s^T \left\{ \hat{J}_t^\top \lambda_t + \frac{1}{2} \text{tr} \left( \Sigma_t^\top \Delta_t^\top \hat{H}_t \Delta_t \Sigma_t \right) \right\} dt \right] = 0. \quad (6)$$

Now we re-write (4) into two terms, the first term $\alpha_t$ containing parameters that are independent of $\Sigma_t \Sigma_t^\top$ and the second (trace) term containing the rest. For instance, $\mu_t$ depends

\[\text{[Footnote: For brevity of notation, we omit dependence of } \hat{J} \text{ and } \hat{H} \text{ on } r \text{ and } s \text{ because we shall see later that it is only the dependence on } t \text{ which is relevant for our proof.]}\]
on $\Sigma_t \Sigma_t^T$ when $x_t$ is a log martingale, but it is zero when $x_t$ is a martingale. The reason for this new decomposition of the drift term in (3) is that we want to choose the parameters in a way that facilitates the derivation of necessary conditions for $(f; u)$ to satisfy (AP). Hence, write

$$\lambda_t = \alpha_t + \frac{1}{2} \left( \text{tr} \left\{ \Sigma_t^T B_t \Sigma_t \right\}, \ldots, \text{tr} \left\{ \Sigma_t^T B_{nt} \Sigma_t \right\} \right)^T,$$  \hspace{1cm} (7)

where $\alpha_t = \alpha (t, x_t) \in \mathbb{R}^n$ and we can assume w.l.o.g. that $B_{it} = B_t (t, x_t) \in \mathbb{R}^{n \times n}$ are symmetric matrices, for $i \in \{1, \ldots, n\}$. \footnote{This is because they only appear within the trace operator, and if $B_{it}$ were not symmetric we can always find another, symmetric matrix with the same trace as $\Sigma_t^T B_{it} \Sigma_t$.}

Next we show that non-trivial solutions arise iff $\alpha = 0$. That is, the only non-zero drift $\lambda_t$ that yields non-trivial solutions must depend on $\Sigma_t \Sigma_t^T$, as would e.g. be the case when $u$ is a log martingale. To see this, insert (7) in (6), to obtain

$$\mathbb{E} \left[ \int_s^T \left\{ \hat{J}_t^\top \alpha_t + \frac{1}{2} \text{tr} \left( \Sigma_t^T \left\{ \hat{J}_1 \Sigma_t B_{1t} + \ldots + \hat{J}_{nt} \Sigma_t B_{nt} + \Delta_t^\top \hat{H}_t \Delta_t \right\} \right) \right\} dt \right] = 0, \hspace{1cm} (8)$$

where $\hat{J} = \left( \hat{J}_1, \ldots, \hat{J}_n \right)^T$. Now consider the spectral decomposition

$$\hat{J}_1 \Sigma_t B_{1t} + \ldots + \hat{J}_{nt} \Sigma_t B_{nt} + \Delta_t^\top \hat{H}_t \Delta_t = W_t V_t W_t^\top,$$  \hspace{1cm} (9)

where $V_t \in \mathbb{R}^{n \times n}$ is a diagonal matrix of eigenvalues and the columns of $W_t \in \mathbb{R}^{n \times n}$ contain the corresponding eigenvectors, which are orthogonal by definition and hence $W_t^\top W_t = I$ (matrix identity). Now we set a specific drift and a specific volatility in our diffusion. For the volatility, we suppose that

$$\Sigma_t = \sigma \exp \left( \frac{\xi}{2} W_t V_t W_t^\top \right) = \sigma W_t \exp \left( \frac{\xi}{2} V_t \right) W_t^\top,$$  \hspace{1cm} (10)

for some real $\xi$ and $\sigma \geq 0$. For the drift we assume

$$\alpha_t = \alpha \hat{J}_t,$$  \hspace{1cm} (11)
for some $\alpha \in \mathbb{R}$. Inserting (9), (10) and (11) into (8), differentiating w.r.t. $T$ and $\xi$ using Leibnitz’ rule for differentiation, and applying the cyclic property of the trace operator yields

$$\alpha E_r \left[ \mathbf{J}_t^\top \mathbf{J}_t \right] = 0, \text{ and } \sigma^2 E_r \left[ \text{tr} (\mathbf{V}_t^2 \exp \{\xi \mathbf{V}_t\}) \right] = 0.$$  

Note that each term in $\mathbf{J}_t^\top \mathbf{J}_t$ is $\geq 0$. Hence, $\alpha E_r \left[ \mathbf{J}_t^\top \mathbf{J}_t \right] = 0 \Rightarrow \alpha = 0$ or $\mathbf{J}_t = \mathbf{0}$, or both; Similarly, each term in $\text{tr} (\mathbf{V}_t^2 \exp \{\xi \mathbf{V}_t\})$ is $\geq 0$. Hence, either $\sigma = 0$ or $\mathbf{V}_t = \mathbf{0}$, or both.

Setting $\sigma = 0$ corresponds to a deterministic process, and $\mathbf{J}_t = \mathbf{0}$ yields the trivial solution $f(u_r, u_s) = a(u_s) - a(u_r)$, since $f(u, u) = 0$ for all $u \in \mathbb{R}^n$. So we assume that $\alpha = 0$ and $\sigma > 0$, which implies that $\alpha_t = 0$ and $\mathbf{V}_t = \mathbf{0}$, for all $t$.

First we show that the condition $\mathbf{V}_t = \mathbf{0}$ implies [PDE I]. To see this, insert $\mathbf{V}_t = \mathbf{0}$ in (9) to yield:

$$\dot{\mathbf{J}}_t \mathbf{B}_t + \ldots + \dot{\mathbf{J}}_n \mathbf{B}_nt + \Delta_t^\top \dot{\mathbf{H}}_t \Delta_t = 0. \quad (12)$$

Rearranging (12), and setting $\mathbf{C}_{it} = - \left( \Delta_t^\top \right)^{-1} \mathbf{B}_{it} \Delta_t^{-1}$, which is symmetric by definition, yields

$$\mathbf{H}_t = - \dot{\mathbf{J}}_t \mathbf{C}_{1t} - \ldots - \dot{\mathbf{J}}_n \mathbf{C}_{nt}. \quad (13)$$

Expanding $\mathbf{H}$ and $\mathbf{J}$ yields

$$\mathbf{H}(u_r, u_t) - \sum_{i=1}^{n} J_i (u_r, u_t) \mathbf{C}_{it} = \mathbf{H}(u_s, u_t) - \sum_{i=1}^{n} J_i (u_s, u_t) \mathbf{C}_{it}. \quad (14)$$

In other words, the expression on the right (and the left) depends only on $u_t$, not on $u_r$ or $u_s$. Therefore there must be some function $a \in \mathcal{C}^2$ which satisfies

$$\mathbf{H}(u_r, u_t) - \sum_{i=1}^{n} J_i (u_r, u_t) \mathbf{C}_{it} = \mathbf{H}^a(u_t) - \sum_{i=1}^{n} J_i^a (u_t) \mathbf{C}_{it}. \quad (15)$$

Then [PDE I] follows on setting $t = s$:

$$\mathbf{H}(u_r, u_s) - \mathbf{H}^a(u_s) = \sum_{i=1}^{n} \{J_i (u_r, u_s) - J_i^a (u_s)\} \mathbf{C}_{is}. \quad (16)$$

24
Note that $x_t$ can be expressed as a function of $u_t$ if $\Delta_t$ is invertible, and the l.h.s. of the above equation depends on $t$ only through $u_t$, hence $C_{it} = C_i(u_t)$. Finally, setting $\alpha_t = 0$ in (7), substituting $B_{it} = -\Delta_t^\top C_{it} \Delta_t$, and equating this with (4), yields (PDE II).

Proof of Corollary 1: Define

$$y_s = u(s,x_s) + \frac{1}{2} \int_t^s \left( \text{tr} \left\{ \Sigma_\tau \Delta_\tau^\top C_{1\tau} \Delta_\tau \Sigma_\tau \right\}, \ldots, \text{tr} \left\{ \Sigma_\tau \Delta_\tau^\top C_{n\tau} \Delta_\tau \Sigma_\tau \right\} \right)^\top d\tau,$$

so that $y_t = u(t,x_t)$. By Itô’s lemma,

$$dy_s = du_s + \frac{1}{2} \left( \text{tr} \left\{ \Sigma_s \Delta_s^\top C_{1s} \Delta_s \Sigma_s \right\}, \ldots, \text{tr} \left\{ \Sigma_s \Delta_s^\top C_{ns} \Delta_s \Sigma_s \right\} \right)^\top ds.$$

Substituting for $du_s$ using (3), and (7) with $\alpha_s = 0$, $\forall s$, yields $dy_s = \Delta_s \Sigma_s dw_s$, i.e. $y_s$ is a martingale. Hence

$$y_t = \mathbb{E}_t [y_T] = \mathbb{E}_t \left[ \psi(x_T) + \frac{1}{2} \int_t^T \left( \text{tr} \left\{ \Sigma_\tau \Delta_\tau^\top C_{1\tau} \Delta_\tau \Sigma_\tau \right\}, \ldots, \text{tr} \left\{ \Sigma_\tau \Delta_\tau^\top C_{n\tau} \Delta_\tau \Sigma_\tau \right\} \right)^\top d\tau \right]$$

$$= u(t,x_t),$$

where we have substituted the boundary condition $u(T,x_T) = \psi(x_T).$

The proof of Theorem 2 requires the following:

Lemma: Let $C_i = QD_iQ^\top$ and $\sum_{i=1}^n C_i = (C_1 1, \ldots, C_n 1)$. Then also

$$\sum_{i=1}^n C_i k_i = (C_1 k, \ldots, C_n k) = (C_1^\top k, \ldots, C_n^\top k)^\top,$$

for all $k = (k_1, \ldots, k_n)^\top \in \mathbb{R}^n$. That is, the third order tensor $(C_1, \ldots, C_n)$ is symmetric.

Proof of Lemma: Let $Q = (q_1, \ldots, q_n)^\top$ and $D_i = \text{diag} (Dq_i)$. Clearly $D$ exists, and it is
unique if $C_i$ are linearly independent. Combining this with $\sum_{i=1}^n C_i = (C_1, \ldots, C_n)$ yields

$$\sum_{i=1}^n Q \text{diag}(Dq_i) Q^\top = (Q \text{diag}(Dq_1) Q^\top 1, \ldots, Q \text{diag}(Dq_n) Q^\top 1).$$

Since $\text{diag}(Dq_i) Q^\top 1 = \text{diag}(Q^\top 1) Dq_i$, we have

$$D = \sum_{i=1}^n \text{diag}(Dq_i) \text{diag}(Q^\top 1)^{-1} = \text{diag}(D1).$$

Therefore $D$ must be diagonal. Then

$$\sum_{i=1}^n C_i k_i = \sum_{i=1}^n Q \text{diag}(Dq_i) Q_i^\top k_i = Q \text{diag}(DQ^\top k) (\text{diag}\{q_1\} 1, \ldots, \text{diag}\{q_n\} 1)
\quad = (Q \text{diag}(Dq_1) Q^\top k, \ldots, Q \text{diag}(Dq_n) Q^\top k) = (C_1k, \ldots, C_nk),$$

where we have used that $\text{diag}(DQ^\top k) \text{diag}(q_i) = \text{diag}(Dq_i) \text{diag}\{Q^\top k\}$. Finally, since $C_i$ are symmetric, we have $(C_1k, \ldots, C_nk) = (C_1^\top k, \ldots, C_n^\top k)^\top$.

**Proof of Theorem 2:** First note that symmetric $C_{it}$ commute for all $i$ and $t$ iff $C_{it} = QD_{it}Q^\top$. See Horn and Johnson [1985], p.52. Then we have first and second partial derivatives of $m$ w.r.t. components of $u$:

$$J^m(u_t) = \exp \left\{ \sum_{i=1}^n \int_0^{u_{it}} C_i(u) \, du_i \right\} = \prod_{i=1}^n \exp \left\{ \int_0^{u_{it}} C_i(u) \, du_i \right\},$$

and $H^m_i(u_t) = C_{it}J^m(u_t) = J^m(u_t)C_{it}$. Now consider

$$\left( \frac{\partial}{\partial u_{1s}}, \ldots, \frac{\partial}{\partial u_{ns}} \right)^\top \left( \{J(u_r, u_s) - J^0(u_s)\}^\top J^m(u_s)^{-1} \right)
= \left( H(u_r, u_s) - H^0(u_s) - \{J(u_r, u_s) - J^0(u_s)\}^\top (C_{1s}, \ldots, C_{ns}) \right) J^m(u_s)^{-1}
= \left( H(u_r, u_s) - H^0(u_s) - \sum_{i=1}^n \{J_i(u_r, u_s) - J^0_i(u_s)\} C_{is} \right) J^m(u_s)^{-1} = 0,
where we apply the Lemma in the second line and \([\text{PDE I}]\) in the third line, and hence

\[
\{J(u_r, u_s) - J^a(u_s)\}^\top J^m(u_s)^{-1} = b(u_r)^\top,
\]

for some differentiable \(b : \mathbb{R}^n \to \mathbb{R}^n\) (note that \(b(u) = J(u, u) - J^a(u)\)). Integration yields

\[
f(u_r, u_s) = a(u_s) - a(u_r) + b(u_r)^\top \{m(u_s) - m(u_r)\},
\]

where we have used that \(f(u, u) = 0\). We can assume w.l.o.g. that \(a(0) = 0\). Furthermore Itô's formula applied to \(m\) yields

\[
dm(u_t) = J^m(u_t) du_t + \frac{1}{2}J^m(u_t) (\text{tr} \{C_1 d \langle u_t \rangle \}, \ldots, \text{tr} \{C_n d \langle u_t \rangle \})^\top = J^m(u_t) \Delta \Sigma d\mathbf{w}_t,
\]

and therefore \(m\) is a multivariate martingale. 

\[\square\]

**Proof of Corollary 2:** Note that

\[
m(u_t) = \int_0^{u_t} \exp \left\{ \sum_{i=1}^{n} C_i u_i \right\} du = \left( \sum_{i=1}^{n} C_i k_i \right)^{-1} \left( \exp \left\{ \sum_{i=1}^{n} C_i u_i \right\} - I \right) k,
\]

for some \(k\) s.t. \(\sum_{i=1}^{n} C_i k_i\) is invertible. First we assume w.l.o.g. that \(k = 1\) and therefore

\[
m(u_T) = \left( \sum_{i=1}^{n} C_i \right)^{-1} \left( \exp \left\{ \sum_{i=1}^{n} C_i \psi_i (x_T) \right\} - I \right) 1.
\]

Now \(m\) is a martingale and we have \(m(u_t) = \mathbb{E}_t [m(u_T)]\), i.e.

\[
m(u_t) = \left( \sum_{i=1}^{n} C_i k_i \right)^{-1} \mathbb{E}_t \left[ \exp \left\{ \sum_{i=1}^{n} C_i \psi_i (x_T) \right\} - 1 \right] 1.
\]

Secondly, since \(\sum_{i=1}^{n} C_i k_i m = \sum_{i=1}^{n} C_i m_i k\) by the Lemma, we can write \(u\) as a function of
rather than the converse, using the equivalent expression to (13), i.e.

\[ u_t = \left( \sum_{i=1}^{n} C_i \right)^{-1} \ln \left( I + \sum_{i=1}^{n} C_i m_{it} \right) \mathbf{1}, \]  

(15)

where the choice of \( \mathbf{1} \) is again arbitrary. Finally, inserting (14) in (15) and once again making use of the Lemma yields

\[ u(t, x_t) = \left( \sum_{i=1}^{n} C_i \right)^{-1} \ln \left( \mathbb{E}_t \left[ \exp \left\{ \sum_{i=1}^{n} C_i \psi_i (x_T) \right\} \right] \right) \mathbf{1}, \]

the log of a martingale satisfying the boundary condition \( u_T = \psi(x_T) \).

\[ \square \]

**Proof of Corollary 3:** We can either derive (M) as the limit for \( C_i \to 0 \) in Theorem 2 or, alternatively, directly from (PDE I) in Theorem 1. For the former derivation, consider a first-order Taylor expansion of the solution

\[ f(u_r, u_s) = a(u_s) - a(u_r) + b(u_r)^\top \left( \sum_{i=1}^{n} C_i \right)^{-1} \sum_{i=1}^{n} C_i (u_{is} - u_{ir}) \mathbf{1} + O \left( \left| \sum_{i=1}^{n} C_i \right| \right), \]

where \( \sum_{i=1}^{n} C_i \) cancels out in the \( b \) term (after applying the Lemma) and the higher orders vanish as \( C_i \to 0, i = 1, \ldots, n \). For the latter derivation, note that setting all \( C_{is} = 0 \) in Theorem 1 yields \( H(u_r, u_s) = H^a(u_s) \). Then

\[ \int_{u_r}^{u_s} \left[ \int_{u_r}^{u_s} d\mathbf{u}_s^\top H(u_r, u_s) \right] d\mathbf{u}_t = \int_{u_r}^{u_s} \left[ \int_{u_r}^{u_s} d\mathbf{u}_s^\top H^a(u_s) \right] d\mathbf{u}_t, \]

and integrating yields

\[ \int_{u_r}^{u_s} [J(u_r, u_t) - J(u_r, u_r)]^\top d\mathbf{u}_t = \int_{u_r}^{u_s} [J^a(u_r) - J^a(u_r)]^\top d\mathbf{u}_t. \]
Integrating the above once again and using \( f(\mathbf{u}, \mathbf{u}) = 0 \) yields the solution

\[
f(\mathbf{u}_r, \mathbf{u}_s) = a(\mathbf{u}_s) - a(\mathbf{u}_r) + b(\mathbf{u}_r)^T (\mathbf{u}_s - \mathbf{u}_r),
\]

where \( b(\mathbf{u}) = J(\mathbf{u}, \mathbf{u}) - J^a(\mathbf{u}) \), so \( b : \mathbb{R}^n \to \mathbb{R}^n \) must be differentiable. Finally, setting \( C_{it} = 0 \) for \( i = 1, \ldots, n \) in Corollary 1 yields \( \mathbf{u}_t = \mathbb{E}_t[\psi(\mathbf{x}_T)] \).

\[\square\]

**Proof of Theorem 3:** For \( t \in \{t_0, \ldots, t_N\} \) we are interested in minimising the expression

\[
\mathbb{E}_t \left[ \left\{ \sum_{i=1}^{N} f(\mathbf{u}_{i-1}, \mathbf{u}_i) - \mathbb{E}_t \left[ \sum_{i=1}^{N} f(\mathbf{u}_{i-1}, \mathbf{u}_i) \right] \right\}^2 \right]
\]

\[
= \mathbb{E}_t \left[ \left\{ a(\mathbf{u}_T) - \mathbb{E}_t [a(\mathbf{u}_T)] + \sum_{i_t > t} b(\mathbf{u}_{i-1})^T (\mathbf{u}_i - \mathbf{u}_{i-1}) \right\}^2 \right]
\]

\[
= \mathbb{E}_t [a(\mathbf{u}_T)^2] - \mathbb{E}_t [a(\mathbf{u}_T)]^2 + 2 \sum_{i_t > t} \mathbb{E}_t \left[ a(\mathbf{u}_T) b(\mathbf{u}_{i-1})^T (\mathbf{u}_i - \mathbf{u}_{i-1}) \right]
\]

\[
+ \sum_{i_t > t} \mathbb{E}_t \left[ \left\{ b(\mathbf{u}_{i-1})^T (\mathbf{u}_i - \mathbf{u}_{i-1}) \right\}^2 \right]
\]

\[
= \mathbb{E}_t [a(\mathbf{u}_T)^2] - \mathbb{E}_t [a(\mathbf{u}_T)]^2 + 2 \sum_{i_t > t} \mathbb{E}_t \left[ b(\mathbf{u}_{i-1})^T \mathbb{E}_{t_{i-1}} [a(\mathbf{u}_T)] (\mathbf{u}_i - \mathbf{u}_{i-1}) \right]
\]

\[
+ \sum_{i_t > t} \mathbb{E}_t \left[ b(\mathbf{u}_{i-1})^T \mathbb{E}_{t_{i-1}} [ (\mathbf{u}_i - \mathbf{u}_{i-1}) (\mathbf{u}_i - \mathbf{u}_{i-1})^T ] b(\mathbf{u}_{i-1}) \right]
\]

\[
= \mathbb{E}_t [a(\mathbf{u}_T)^2] - \mathbb{E}_t [a(\mathbf{u}_T)]^2 + \sum_{i_t > t} \mathbb{E}_t \left[ 2b(\mathbf{u}_{i-1})^T \omega_{i-1} + b(\mathbf{u}_{i-1})^T \Omega_{i-1} b(\mathbf{u}_{i-1}) \right],
\]

with \( \omega_{i-1} = \mathbb{E}_{t_{i-1}} [a(\mathbf{u}_T)] (\mathbf{u}_i - \mathbf{u}_{i-1}) \) and \( \Omega_{i-1} = \mathbb{E}_{t_{i-1}} \left[ (\mathbf{u}_i - \mathbf{u}_{i-1}) (\mathbf{u}_i - \mathbf{u}_{i-1})^T \right] \). Taking the derivative with respect to the components of \( b(\mathbf{u}_T) \) and setting the result equal to zero yields \( b(\mathbf{u}_T) = -\Omega_{i-1}^{-1} \omega_t \). The second derivative w.r.t. \( b \) corresponds to \( \Omega \) and is positive definite as long as the components of \( \psi \) are linearly independent. We have therefore found a unique minimum.

Assume now that \( \mathbf{u} \) follows the dynamics \( d\mathbf{u}_t = \mathbf{\Delta}_t \Sigma_t dw_t \) and, for some \( r < s \), consider \( \omega_r = \mathbb{E}_r \left[ (\mathbb{E}_t [a(\mathbf{u}_T)] - \mathbb{E}_r [a(\mathbf{u}_T)]) (\mathbf{u}_s - \mathbf{u}_r) \right] \) and \( \Omega_r = \mathbb{E}_r \left[ (\mathbf{u}_s - \mathbf{u}_r) (\mathbf{u}_s - \mathbf{u}_r)^T \right] \). Then, as \( r \to s \) (continuous monitoring), we have \( \Omega_t = d \langle \mathbf{u} \rangle_t \) and \( \omega_t = d \langle \mathbf{u} \rangle_t J^a(\mathbf{u}_t) \) (using that
\[ d \mathbb{E}_t [a(u_t)] = J^a (u_t)^\top d u_t , \] and therefore \[ b(u) = - \Omega^{-1} \omega = - J^a (u). \]

Alternatively, the same result can be obtained from first-order Taylor expansion. To see this, consider \[ f(u, u + \delta u) = \{J^a (u) + b(u)\}^\top \delta u + O (|\delta u|^2) , \] and note that the leading order of \( \delta u \) is zero for \( b(u) = - J^a (u) \).

\[ \square \]

**Proof of Corollary 4:** Inserting the specifications of \( u, a \) and \( b \) in \([M]\) yields

\[
f(u_r, u_s) = \left( c_5 + 2c_6^\lambda - 2c_6^\eta \right) (\ln F_s - \ln F_r) + 4c_8 \left( Y_s^2 - Y_r^2 \right) + 2c_6^\eta \left( \frac{Z_s}{F_s} - \frac{Z_r}{F_r} \right) \\
+ \left( \frac{c_7}{F_r} - \frac{2c_9 z_r}{F_r^2} \right) (F_s - F_r) + \left( -2c_6^\lambda - 8c_8 Y_r \right) (Y_s - Y_r) + 2c_9 \left( Z_s - Z_r \right) \\
+ c_7 e^{-y_r} (e^{y_s} - e^{y_r}) + 4c_8 \left( Y_s^2 - 2Y_r (Y_s - Y_r) - Y_r^2 \right) \\
+ 2c_9 \left( Z_s - Z_r - \frac{Z_r}{F_r} (F_s - F_r) \right) \\
= c_5 (y_s - y_r) + c_6^\lambda (v_s^\lambda - v_r^\lambda) + c_6^\eta (v_s^\eta - v_r^\eta) + c_7 (e^{y_s - y_r} - 1) \\
+ c_8 \left( 2Y_s - 2Y_r \right)^2 + c_9 \left( \frac{2Z_s}{F_s} - \frac{2Z_r}{F_r} \right) e^{y_s - y_r} = g(x_s - x_r) . \]

\[ \square \]