A general property for time aggregation

Article (Accepted Version)

Alexander, Carol and Rauch, Johannes (2021) A general property for time aggregation. *European Journal of Operational Research*, 291 (2). pp. 536-548. ISSN 0377-2217

This version is available from Sussex Research Online: <http://sro.sussex.ac.uk/id/eprint/89186/>

This document is made available in accordance with publisher policies and may differ from the published version or from the version of record. If you wish to cite this item you are advised to consult the publisher's version. Please see the URL above for details on accessing the published version.

Copyright and reuse:

Sussex Research Online is a digital repository of the research output of the University.

Copyright and all moral rights to the version of the paper presented here belong to the individual author(s) and/or other copyright owners. To the extent reasonable and practicable, the material made available in SRO has been checked for eligibility before being made available.

Copies of full text items generally can be reproduced, displayed or performed and given to third parties in any format or medium for personal research or study, educational, or not-for-profit purposes without prior permission or charge, provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way.

A General Property for Time Aggregation

Carol Alexander and Johannes Rauch*

Accepted for publication in European Journal of Operational Research on 3 January 2020

Abstract

We classify all functions of multivariate stochastic processes having time-series estimates that are independent of data frequency. Such an estimator applied to high-frequency data may be used to infer properties of estimates relating to low-frequency data. Our property encompasses two previously-proposed time-aggregation properties (with limited solutions) as different special cases. Our general time-aggregating functions satisfy a pair of coupled second-order partial differential equations. We derive analytic solutions for arbitrary-dimensional martingales and log-martingales. The time-aggregation property of a time-series model is similar – indeed time-aggregating functions always correspond to point estimators based on expected values – but we do not propose a specific new forecasting model. However, we do derive time-aggregating unbiased and efficient estimators for n^{th} -order moments of log returns, applying these results to problems facing portfolio managers who re-optimize portfolios or hedge their risks at lower frequencies than the frequency at which their risk premia are monitored.

Keywords: : Aggregation Property; Higher Moments; Risk Premia; Time Aggregation; Unbiased and Efficient Estimators

JEL Classification: C10; G10

*University of Sussex Business School, UK. Corresponding Author email: c.alexander@sussex.ac.uk

The classical temporal aggregation property refers to a discrete time-series model whose structure is invariant to time partitions of the stochastic process on which it is defined. For instance, consider the standard autoregressive moving average model applied to the standard stochastic process in finance, i.e. the log return of a financial asset price.¹ The model satisfies the temporal aggregation property for this process if it retains the same structure whether the log returns are sampled annually, monthly, weekly, daily, or hourly. Put another way, when realisations are generated from a time-aggregating model then samples of aggregated realisations still have the same model structure. The substantial literature on this topic is thoroughly surveyed by Silvestrini and Veredas [2008] and more recently by Mamingi [2017]. Almost all papers examine linear models within the general vector autoregressive integrated moving average class, see Tsai and Chan [2005], Foroni et al. [2018] and many others.

A general application of the temporal aggregation property is to improve point forecasts for a stochastic process – e.g. to use low-frequency data to generate short-term forecasts, or high-frequency data for long-term forecasts. This question has already been thoroughly explored for linear models in financial econometrics. See Taylor [2008], Hafner [2009] and Kole et al. [2017] amongst many others. But non-linear models are more difficult to handle and very few attempts to derive aggregation results have been successful.² Much of the data used in finance and economics shows strong evidence of non-linearity in its structure, so theoretical results on temporal aggregation for non-linear models are particularly important for such applications [Aghababa and Barnett, 2016]. The temporal aggregation property has also found specific applications to many areas of operational research, including hydrology [Forsman et al., 2003, Lorite et al., 2005, Bastola and Misra, 2013] and to transportation/demand problems in supply chain management [Jin et al., 2015, Athanasopoulos et al., 2017, Rostami-Tabar et al., 2019]. A topic in time-series forecasting closely related to temporal aggregation is combining predictions based on different frequency data with the aim of producing superior forecasts. Starting with the work of Hyndman et al. [2011] and subsequently developed by Kourentzes et al. [2014], Petropoulos and Kourentzes [2015] and Kourentzes and Petropoulos [2016] a productive stand of recent research explores the performance of forecasts based on the concept of

¹This is the difference between the log of the asset price S at two different points in time: the backward-looking h -period log return at time t is $\ln S_t - \ln S_{t-h}$ is for estimation and the forward-looking log return at time t is $\ln S_{t+h} - \ln S_t$ is for forecasting.

²A notable exception is Granger [2008], who shows that any non-linear model can be approximated by a time-varying parameter linear model, whence approximate (and fairly complex) theoretical results on aggregation are obtained.

“multiple temporal hierarchies” which is formalised by Athanasopoulos et al. [2017].

By contrast, this paper does not develop a forecasting model. The motivation for our research stems from an aggregation property *only* in expectations, not for the entire model structure, which has its roots in the finance literature on the variance risk premium (VRP). This premium is defined as the difference between a realised variance (RV) based on time series of log returns (in the physical measure) and the variance that is implied from current prices of standard European options (in the risk-neutral measure). The VRP may be identified with the pay-off to a financial contract called a variance swap,³ and in order to derive error-free estimates and forecasts for the VRP both Neuberger [2012] and Bondarenko [2014] introduce (non-equivalent) aggregation properties for the expected value of a generalised variance process.

A point forecast is generated by taking expectations of a fitted model. If the temporally aggregated model is only used to generate improved forecasts, then it is the temporal aggregation property of expectations that really matters. The question whether the model structure itself is invariant to the partition is less relevant than the question of whether point forecasts are the same, irrespective of the data frequency used to generate them. Put another way, what matters is whether we obtain the same expected value from high frequency data as we would from low frequency data.

In the realm of linear models the answer is trivial – all expectations of linear functions have this aggregation property. Indeed, the additivity of the expectation operator explains the prevalence of linear models in the time-series literature. But our paper asks a much more difficult question: which non-linear functions of which stochastic processes have the temporal aggregation property in expectations?

Partial answers to this question have already been addressed by Neuberger [2012] and Bondarenko [2014], who derive (different) specific solutions to what they both term the ‘aggregation property’ for non-linear functions of stochastic processes. However, each solution is limited to a different special case, described in the next section. The advantages of our results are that we derive: (1) the *general* properties that must be satisfied by any aggregating function of any multivariate

³Assuming (hypothetically) that monitoring of the RV is continuous and that there are no jumps in the forward price, a unique fair-value variance swap rate is derived from market prices of traded options using the Carr and Madan [2001] replication integral. However, in the real world neither of these assumptions hold and many papers investigate the discretisation error: [Broadie and Jain, 2008, Carr and Wu, 2009, Carr and Lee, 2009, Davis et al., 2014, Hobson and Klimmek, 2012, Jarrow et al., 2013, Bernard and Cui, 2014, Bernard et al., 2014]; and jump error [Ait-Sahalia, 2004, Rompolis and Tzavalis, 2017] and many others. The combination of these errors lead to large deviations between fair-value and market swap rates. For instance, Ait-Sahalia et al. [2018] show that variance swap rates for the S&P 500 are frequently 5% or more above the fair value determined by the CBOE Volatility Index (VIX).

stochastic processes; and (2) analytic solutions for *any* multivariate martingale or log-martingale process. We also show that our general solution encompass both the Neuberger [2012] and Bondarenko [2014] results, each as a different special case. As an application, we focus on the functions which yield aggregating expectations for higher moments, because these are particularly relevant for the unbiased and efficient estimation of the multi-asset risk premia which are often applied in the variance and skewness swap strands of asset pricing literature.

In the following: Section 1 introduces the aggregation properties of Neuberger [2012] and Bondarenko [2014]; Section 2 presents the theoretical results on our general property; Section 3 derives the aggregating characteristics which correspond to higher moments; Section 4 provides an empirical application using real data, to compare (and swap) higher-moment risk premia corresponding to different frequencies and Section 5 concludes.

1 Two Aggregation Properties

A real-valued function f defined on a time-dependent stochastic process $\mathbf{x}_t$ satisfies the aggregation property (AP) if and only if the expected value of a sum is invariant to the partition of the interval over which it is defined. For example, suppose $\mathbf{x}_t$ is just a forward price F_t for $0 \leq t \leq T$ in an arbitrage-free market. Then F_t is a martingale under some measure. Set $f(F_t, F_s) = (F_t - F_s)^2$. Then a straightforward application of the tower law of conditional expectations shows that the expected value (under that measure) of a sum of squared increments in F_t is independent of the partition for the sum. It does not matter whether the increments are monitored continuously, hourly, daily or weekly over the interval $[0, T]$; indeed, the partition π for the sum doesn't even need to be a regular one, we have:

$$\mathbb{E}_0 \left[\sum_{\pi} (F_{t_i} - F_{t_{i-1}})^2 \right] = \mathbb{E} \left[\sum_{t=1}^T (F_t - F_{t-1})^2 \right] = \mathbb{E}_0 [(F_T - F_0)^2]. \quad (1)$$

However, when F_t is a martingale the standard RV, which is based on a sum of squared *log* returns, does not satisfy the AP. That is, $\mathbb{E}_0 [\sum_{\pi} (\ln F_{t_i} - \ln F_{t_{i-1}})^2] \neq \mathbb{E} [(\ln F_T - \ln F_0)^2]$ for an arbitrary partition π . This is the root of the major biases in the fair-value formula for a standard variance swap mentioned in our introduction, i.e. the discretisation error (the formula assumes the RV is monitored continuously) and the jump error (the integral is derived by assuming F_t follows a diffusion – not necessarily geometric Brownian motion). However, the fair value of

an arithmetic variance swap, for which the realised characteristic is based on (1), has neither jump nor discretisation bias [Leontsinis and Alexander, 2017]. Indeed, *any* swap having a realised characteristic which satisfies the AP has neither discretisation error nor jump error. It will have an unbiased and unique fair value which may be derived using the replication integral of Carr and Madan [2001].⁴

This property was first introduced by Neuberger [2012] and Bondarenko [2014] independently, but it is not possible to write one AP in terms of the other. Furthermore, each is restricted to a narrow domain: Neuberger [2012] only considers functions of increments, albeit for a multivariate stochastic process; and Bondarenko [2014] only considers functions of a univariate process. Both authors focus on functions which correspond to second-order moments of log returns, and each derives solutions which correspond to generalised variance processes. Neuberger [2012] also finds one third-order moment which satisfies his AP. However, thus far it is not clear whether their approaches can be generalised to other functions, or whether the two definitions can be unified as special cases of a more general AP.

Neuberger [2012] defines his AP as follows: given an adapted stochastic process $\mathbf{x}_t$ on a standard filtered probability space and the real-valued function $g(\mathbf{x})$, the pair $(g; \mathbf{x})$ satisfies the aggregation property (APN) iff

$$\mathbb{E}_r [g(\mathbf{x}_T - \mathbf{x}_r)] = \mathbb{E}_r [g(\mathbf{x}_s - \mathbf{x}_r)] + \mathbb{E}_r [g(\mathbf{x}_T - \mathbf{x}_s)] \quad \forall \quad 0 \leq r \leq s \leq T, \quad (\text{APN})$$

where $\mathbb{E}_t$ denotes the expectation under some probability measure conditional on the filtration at time t . Applying the tower law of conditional expectations to (APN) yields

$$\mathbb{E}_0 \left[\sum_{i=1}^N g(\delta \mathbf{x}_i) \right] = \mathbb{E}_0 [g(\mathbf{x}_T - \mathbf{x}_0)],$$

for any partition $\{0 = t_0 < t_1 < \dots < t_N = T\}$, where $\delta \mathbf{x}_i = \mathbf{x}_{t_i} - \mathbf{x}_{t_{i-1}}$. 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 – hence the absence of jump error.

Bondarenko [2014] introduces an alternative version of the aggregation property (APB) based on the levels rather than the increments of a $\mathcal{C}^1$ function $h : \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$, as:

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

He shows that when x is a martingale the solutions to (APB) are given by $h(x_r, x_s) = a(x_s) -$

⁴The only errors in its computation will arise from the need to integrate numerically over the available strikes but these errors are typically quite small [Jiang and Tian, 2005].

$a(x_r) + b(x_r)(x_s - x_r)$ 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.⁵

In the univariate case (APB) is more general than (APN) in the sense that if the pair $(g; x)$ satisfies (APN) then $h(x_r, x_s) = g(x_s - x_r)$ satisfies (APB). It may at first appear that it is possible to write (APB) in terms of (APN), by doubling the size of the state space. That is, setting $\mathbf{x} = (F, y)^\top$, 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 $\delta F, \delta y \rightarrow 0$ even when h is a polynomial or some other well-behaved function. Hence, (APN) does not include (APB).

This paper extends the definitions of Neuberger [2012] and Bondarenko [2014] to a general AP which encompasses each as a different special case. The innovation which allows this is to consider functions f of a ‘*derivative*’ process $\mathbf{u}_t = \mathbf{u}(t, \mathbf{x}_t)$ of the multivariate process $\mathbf{x}_t$. Assuming only that second-order derivatives exist for both f and $\mathbf{u}$, a pair of second order systems of partial differential equations form necessary conditions for $(f; \mathbf{u})$ to satisfy our AP. When $\mathbf{u}_t$ contains only log martingales, or when it contains only martingales, we can derive all solutions to these equations in closed form. This way, we classify an entire vector space of functions which satisfy the AP. Thereafter we focus on moment characteristics, i.e. those solutions of AP which correspond to unbiased estimates for higher moments of log returns of a martingale forward price. There are still an infinite number of solutions, but we show how to select the most efficient among them. Then we compare our unbiased and efficient variance and third-moment characteristics with the two specific AP solutions found by Neuberger [2012], and we also explore the properties of an unbiased and efficient fourth-moment characteristic.

2 The General Aggregation Property

Let $\mathbf{x}_t \in \mathbb{R}^n$ for $t \in [0, T]$ denote a multivariate adapted process on a standard filtered probability space and consider the derivative process $\mathbf{u}_t = \mathbf{u}(t, \mathbf{x}_t)$ for a vector-valued function $\mathbf{u} : [0, T] \times \mathbb{R}^n \rightarrow \mathbb{R}^n$ as well as a real-valued function $f : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}$. This setting allows the unification of the aggregation properties of Neuberger [2012] and Bondarenko [2014] in the single definition as follows:

⁵Bondarenko [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 [2019], Equation 4, in the context of realised divergence.

$$\mathbb{E}_r[f(\mathbf{u}_r, \mathbf{u}_T)] = \mathbb{E}_r[f(\mathbf{u}_r, \mathbf{u}_s)] + \mathbb{E}_r[f(\mathbf{u}_s, \mathbf{u}_T)] \quad \forall \quad 0 \leq r \leq s \leq T. \quad (\text{AP})$$

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

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

$$\mathbb{E}_0 \left[\sum_{i=1}^N f(\mathbf{u}_{i-1}, \mathbf{u}_i) \right] = \mathbb{E}_0[f(\mathbf{u}_0, \mathbf{u}_T)], \quad (\text{E})$$

for all partitions $\{0 = t_0 < t_1 < \dots < t_N = T\}$, where we write $\mathbf{u}_i = \mathbf{u}_{t_i}$ for convenience. As in Neuberger [2012], the interpretation of (E) depends on the measure: if (AP) holds under the physical measure, then $\sum_{i=1}^N f(\mathbf{u}_{i-1}, \mathbf{u}_i)$ is an unbiased estimator of $\mathbb{E}_0[f(\mathbf{u}_0, \mathbf{u}_T)]$ for any partition of $[0, T]$. If (AP) holds under a pricing measure, then the fair value of a contingent claim that pays $\sum_{i=1}^N f(\mathbf{u}_{i-1}, \mathbf{u}_i)$ is the same as that of a contingent claim which pays $f(\mathbf{u}_0, \mathbf{u}_T)$. Under the additional assumption that this contingent claim exists or can be synthesised from other claims, a fair value can be derived from the market, and because the r.h.s. of (E) is path independent, this fair value has neither jump nor discretisation error.

Our (AP) is more general than (APB) in that we consider an n -dimensional multivariate derivative process which may contain elements that are either martingales or not; and more general than (APN) because our function is defined on the levels at the start and end of an interval in the partition, rather than increments δF over successive intervals. Clearly (APN) is the special case of (AP) where $f(\mathbf{u}_r, \mathbf{u}_s) = g(\mathbf{x}_s - \mathbf{x}_r)$.

Our first result characterises the pairs $(f; \mathbf{u})$, with $f : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}$, which satisfy (AP) as solutions to two second-order partial differential equations, one for the derivative processes $\mathbf{u} = (u_1, \dots, u_n)^\top$ w.r.t. the underlying processes $\mathbf{x} = (x_1, \dots, x_n)^\top$ and the other for the real-valued function f w.r.t. $\mathbf{u}$. We only need to make some minimal assumptions on f and $\mathbf{u}$, i.e. that the following quantities exist: For $i \in \{1, \dots, n\}$ let $\boldsymbol{\vartheta}_t = (\vartheta_{1t}, \dots, \vartheta_{nt})^\top$ where $\vartheta_{it} = \vartheta_i(t, \mathbf{x}_t) \in \mathbb{R}$ denote the time derivatives of $\mathbf{u}$, and let $\boldsymbol{\delta}_{it} = \boldsymbol{\delta}_i(t, \mathbf{x}_t) \in \mathbb{R}^n$ and $\boldsymbol{\Gamma}_{it} = \boldsymbol{\Gamma}_i(t, \mathbf{x}_t) \in \mathbb{R}^{n \times n}$ denote the first and second partial derivatives of $\mathbf{u}_t = \mathbf{u}(t, \mathbf{x}_t)$ w.r.t. the i^{th} component of $\mathbf{x}_t$. For the function f we denote the Jacobian vector of first partial derivatives w.r.t. the components of the second input vector by $\mathbf{J} = (J_1, \dots, J_n)^\top \in \mathbb{R}^n$ and write $\mathbf{H} \in \mathbb{R}^{n \times n}$ for the Hessian matrix of second partial derivatives. With these definitions we can now establish *necessary* conditions for

$(f; \mathbf{u})$ to satisfy the (AP), by considering a particular process for $\mathbf{x}_t$ (a multivariate diffusion with a particular drift and covariance) and then deriving conditions for the AP to hold.

Theorem 1: Assume that $\mathbf{x}$ follows a diffusion process with dynamics $d\mathbf{x}_t = \boldsymbol{\mu}_t dt + \boldsymbol{\Sigma}_t d\mathbf{w}_t$, where $\boldsymbol{\Sigma}_t = \boldsymbol{\Sigma}(t, \mathbf{x}_t) \in \mathbb{R}^{n \times n}$, $\boldsymbol{\mu}_t = \boldsymbol{\mu}(t, \mathbf{x}_t) \in \mathbb{R}^n$ and $\mathbf{w}_t \in \mathbb{R}^n$ is a standard multivariate Wiener process, for $t \in [0, T]$. If the first derivatives of $\mathbf{u}$ w.r.t. $\mathbf{x}$, at time t , viz. $\boldsymbol{\Delta}_t = (\boldsymbol{\delta}_{1t}, \dots, \boldsymbol{\delta}_{nt})^\top$ form an $n \times n$ invertible matrix then for every $(f; \mathbf{u})$ with $f(\mathbf{u}, \mathbf{u}) = 0 \forall \mathbf{u} \in \mathbb{R}^n$, there exists a function $a \in \mathcal{C}^2$ such that:

$$\mathbf{H}(\mathbf{u}_r, \mathbf{u}_s) - \mathbf{H}^a(\mathbf{u}_s) = \sum_{i=1}^n \{J_i(\mathbf{u}_r, \mathbf{u}_s) - J_i^a(\mathbf{u}_s)\} \mathbf{C}_{is}, \quad (\text{PDE I})$$

for any $0 \leq r \leq s \leq T$, where $\mathbf{J}^a = (J_1^a, \dots, J_n^a)^\top \in \mathbb{R}^n$ and $\mathbf{H}^a \in \mathbb{R}^{n \times n}$ are the Jacobian and Hessian of a and $\mathbf{C}_{it} = \mathbf{C}_i(\mathbf{u}_t) \in \mathbb{R}^{n \times n}$ are symmetric matrices, for $i = 1, \dots, n$. Moreover:

$$2(\boldsymbol{\vartheta}_t + \boldsymbol{\Delta}_t \boldsymbol{\mu}_t) + \left(\text{tr} \left\{ \boldsymbol{\Sigma}_t^\top \mathbf{M}_{1t} \boldsymbol{\Sigma}_t \right\}, \dots, \text{tr} \left\{ \boldsymbol{\Sigma}_t^\top \mathbf{M}_{nt} \boldsymbol{\Sigma}_t \right\} \right)^\top = \mathbf{0}, \quad (\text{PDE II})$$

where $\mathbf{M}_{it} = \boldsymbol{\Gamma}_{it} + \boldsymbol{\Delta}_t^\top \mathbf{C}_{it} \boldsymbol{\Delta}_t$, for all $t \in [0, T]$.

Proof of Theorem 1: Decomposing $f(\mathbf{u}_r, \mathbf{u}_s)$ using Itô integrals yields

$$f(\mathbf{u}_r, \mathbf{u}_s) = \int_r^s \mathbf{J}(\mathbf{u}_r, \mathbf{u}_t)^\top d\mathbf{u}_t + \frac{1}{2} \int_r^s \text{tr} \{ \mathbf{H}(\mathbf{u}_r, \mathbf{u}_t) d\langle \mathbf{u} \rangle_t \}. \quad (2)$$

Applying (2) to all terms in (AP), and writing $\hat{\mathbf{J}}_t = \mathbf{J}(\mathbf{u}_r, \mathbf{u}_t) - \mathbf{J}(\mathbf{u}_s, \mathbf{u}_t)$ as well as $\hat{\mathbf{H}}_t = \mathbf{H}(\mathbf{u}_r, \mathbf{u}_t) - \mathbf{H}(\mathbf{u}_s, \mathbf{u}_t)$,⁶ yields

$$\mathbb{E}_r \left[\int_s^T \hat{\mathbf{J}}_t^\top d\mathbf{u}_t + \frac{1}{2} \int_s^T \text{tr} \left(\hat{\mathbf{H}}_t d\langle \mathbf{u} \rangle_t \right) \right] = 0, \quad (3)$$

Since $d\mathbf{x}_t = \boldsymbol{\mu}_t dt + \boldsymbol{\Sigma}_t d\mathbf{w}_t$, Itô's lemma applied to $\mathbf{u}_t = \mathbf{u}(t, \mathbf{x}_t)$ yields

$$d\mathbf{u}_t = \boldsymbol{\lambda}_t dt + \boldsymbol{\Delta}_t \boldsymbol{\Sigma}_t d\mathbf{w}_t, \quad (4)$$

where

$$\boldsymbol{\lambda}_t = \boldsymbol{\vartheta}_t + \boldsymbol{\Delta}_t \boldsymbol{\mu}_t + \frac{1}{2} \left(\text{tr} \left\{ \boldsymbol{\Sigma}_t^\top \boldsymbol{\Gamma}_{1t} \boldsymbol{\Sigma}_t \right\}, \dots, \text{tr} \left\{ \boldsymbol{\Sigma}_t^\top \boldsymbol{\Gamma}_{nt} \boldsymbol{\Sigma}_t \right\} \right)^\top. \quad (5)$$

The quadratic variation of $\mathbf{x}$ is $d\langle \mathbf{x} \rangle_t = \boldsymbol{\Sigma}_t \boldsymbol{\Sigma}_t^\top dt$, and so

$$d\langle \mathbf{u} \rangle_t = \boldsymbol{\Delta}_t \boldsymbol{\Sigma}_t \boldsymbol{\Sigma}_t^\top \boldsymbol{\Delta}_t^\top dt. \quad (6)$$

Inserting (4), (5) and (6) in (3), and noting that the stochastic integral w.r.t. $d\mathbf{w}$ vanishes under expectation, since $\mathbf{w}$ is a martingale, yields

$$\mathbb{E}_r \left[\int_s^T \left\{ \hat{\mathbf{J}}_t^\top \boldsymbol{\lambda}_t + \frac{1}{2} \text{tr} \left(\boldsymbol{\Sigma}_t^\top \boldsymbol{\Delta}_t^\top \hat{\mathbf{H}}_t \boldsymbol{\Delta}_t \boldsymbol{\Sigma}_t \right) \right\} dt \right] = 0. \quad (7)$$

⁶For brevity of notation, we omit dependence of $\hat{\mathbf{J}}$ and $\hat{\mathbf{H}}$ on r and s because we shall see later that it is only the dependence on t which is relevant for our proof.

Now we re-write (5) into two terms, the first term α_t containing parameters that are independent of $\Sigma_t \Sigma_t^\top$ and the second (trace) term containing the rest. For instance, μ_t depends on $\Sigma_t \Sigma_t^\top$ when $\mathbf{x}_t$ is a log martingale, but it is zero when $\mathbf{x}_t$ is a martingale. The reason for this new decomposition of the drift term in (4) is that we want to choose the parameters in a way that facilitates the derivation of necessary conditions for $(f; \mathbf{u})$ to satisfy (AP). Hence, write

$$\lambda_t = \alpha_t + \frac{1}{2} \left(\text{tr} \left\{ \Sigma_t^\top \mathcal{B}_{1t} \Sigma_t \right\}, \dots, \text{tr} \left\{ \Sigma_t^\top \mathcal{B}_{nt} \Sigma_t \right\} \right)^\top, \quad (8)$$

where $\alpha_t = \alpha(t, \mathbf{x}_t) \in \mathbb{R}^n$ and we can assume w.l.o.g. that $\mathcal{B}_{it} = \mathcal{B}_i(t, \mathbf{x}_t) \in \mathbb{R}^{n \times n}$ are symmetric matrices, for $i \in \{1, \dots, n\}$.⁷

Next we show that non-trivial solutions arise iff $\alpha = \mathbf{0}$. That is, the only non-zero drift λ_t that yields non-trivial solutions must depend on $\Sigma_t \Sigma_t^\top$, as would e.g. be the case when $\mathbf{u}$ is a log martingale. To see this, insert (8) in (7), to obtain

$$\mathbb{E}_r \left[\int_s^T \left\{ \hat{\mathbf{J}}_t^\top \alpha_t + \frac{1}{2} \text{tr} \left(\Sigma_t^\top \left\{ \hat{J}_{1t} \mathcal{B}_{1t} + \dots + \hat{J}_{nt} \mathcal{B}_{nt} + \Delta_t^\top \hat{\mathbf{H}}_t \Delta_t \right\} \Sigma_t \right) \right\} dt \right] = 0, \quad (9)$$

where $\hat{\mathbf{J}} = (\hat{J}_1, \dots, \hat{J}_n)^\top$. Now consider the spectral decomposition

$$\hat{J}_{1t} \mathcal{B}_{1t} + \dots + \hat{J}_{nt} \mathcal{B}_{nt} + \Delta_t^\top \hat{\mathbf{H}}_t \Delta_t = \mathbf{W}_t \mathbf{V}_t \mathbf{W}_t^\top, \quad (10)$$

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

suppose that $\Sigma_t = \sigma \exp \left(\frac{\xi}{2} \mathbf{W}_t \mathbf{V}_t \mathbf{W}_t^\top \right) = \sigma \mathbf{W}_t \exp \left(\frac{\xi}{2} \mathbf{V}_t \right) \mathbf{W}_t^\top$,
for some real ξ and $\sigma \geq 0$. For the drift we assume

$$\alpha_t = \alpha \hat{\mathbf{J}}_t, \quad (12)$$

for some $\alpha \in \mathbb{R}$. Inserting (10), (11) and (12) into (9), differentiating w.r.t. T and ξ using Leibnitz' rule for differentiation, and applying the cyclic property of the trace operator yields

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

Note that each term in $\hat{\mathbf{J}}_t^\top \hat{\mathbf{J}}_t$ is ≥ 0 . Hence, $\alpha \mathbb{E}_r \left[\hat{\mathbf{J}}_t^\top \hat{\mathbf{J}}_t \right] = 0 \Rightarrow \alpha = 0$ or $\hat{\mathbf{J}}_t = \mathbf{0}$, or both; Similarly, each term in $\text{tr} \left(\mathbf{V}_t^2 \exp \{ \xi \mathbf{V}_t \} \right)$ is ≥ 0 . Hence, either $\sigma = 0$ or $\mathbf{V}_t = \mathbf{0}$, or both. Setting $\sigma = 0$ corresponds to a deterministic process, and $\hat{\mathbf{J}}_t = \mathbf{0}$ yields the trivial solution $f(\mathbf{u}_r, \mathbf{u}_s) =$

⁷This is because they only appear within the trace operator, and if $\mathcal{B}_{it}$ were not symmetric we can always find another, symmetric matrix with the same trace as $\Sigma_t^\top \mathcal{B}_{it} \Sigma_t$.

$a(\mathbf{u}_s) - a(\mathbf{u}_r)$, since $f(\mathbf{u}, \mathbf{u}) = 0$ for all $\mathbf{u} \in \mathbb{R}^n$. So we assume that $\alpha = 0$ and $\sigma > 0$, which implies that $\boldsymbol{\alpha}_t = \mathbf{0}$ 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 (10) to yield:

$$\hat{J}_{1t}\mathcal{B}_{1t} + \dots + \hat{J}_{nt}\mathcal{B}_{nt} + \boldsymbol{\Delta}_t^\top \hat{\mathbf{H}}_t \boldsymbol{\Delta}_t = \mathbf{0}. \quad (13)$$

Rearranging (13), and setting $\mathbf{C}_{it} = -(\boldsymbol{\Delta}_t^\top)^{-1} \mathcal{B}_{it} \boldsymbol{\Delta}_t^{-1}$, which is symmetric by definition, yields $\hat{\mathbf{H}}_t = -\hat{J}_{1t}\mathbf{C}_{1t} - \dots - \hat{J}_{nt}\mathbf{C}_{nt}$. Expanding $\hat{\mathbf{H}}$ and $\hat{\mathbf{J}}$ yields

$$\mathbf{H}(\mathbf{u}_r, \mathbf{u}_t) - \sum_{i=1}^n J_i(\mathbf{u}_r, \mathbf{u}_t) \mathbf{C}_{it} = \mathbf{H}(\mathbf{u}_s, \mathbf{u}_t) - \sum_{i=1}^n J_i(\mathbf{u}_s, \mathbf{u}_t) \mathbf{C}_{it}.$$

In other words, the expression on the right (and the left) depends only on $\mathbf{u}_t$, not on $\mathbf{u}_r$ or $\mathbf{u}_s$.

Therefore there must be some function $a \in \mathcal{C}^2$ which satisfies

$$\mathbf{H}(\mathbf{u}_r, \mathbf{u}_t) - \sum_{i=1}^n J_i(\mathbf{u}_r, \mathbf{u}_t) \mathbf{C}_{it} = \mathbf{H}^a(\mathbf{u}_t) - \sum_{i=1}^n J_i^a(\mathbf{u}_t) \mathbf{C}_{it}.$$

Then (PDE I) follows on setting $t = s$:

$$\mathbf{H}(\mathbf{u}_r, \mathbf{u}_s) - \mathbf{H}^a(\mathbf{u}_s) = \sum_{i=1}^n \{J_i(\mathbf{u}_r, \mathbf{u}_s) - J_i^a(\mathbf{u}_s)\} \mathbf{C}_{is}.$$

Note that $\mathbf{x}_t$ can be expressed as a function of $\mathbf{u}_t$ if $\boldsymbol{\Delta}_t$ is invertible, and the l.h.s. of the above equation depends on t only through $\mathbf{u}_t$, hence $\mathbf{C}_{it} = \mathbf{C}_i(\mathbf{u}_t)$. Finally, setting $\boldsymbol{\alpha}_t = \mathbf{0}$ in (8), substituting $\mathcal{B}_{it} = -\boldsymbol{\Delta}_t^\top \mathbf{C}_{it} \boldsymbol{\Delta}_t$, and equating this with (5), yields (PDE II). $\square$

Having derived the necessary conditions (PDE I) and (PDE II) we now provide the solutions to (AP) for symmetric matrices $\mathbf{C}_{it}$ and derive closed-form expressions for $(f; \mathbf{u})$ in two special cases, viz. Corollary 2, when all $\mathbf{C}_{it}$ are constant – so that each process in $\mathbf{u}$ follows a log martingale – and Corollary 3 when all $\mathbf{C}_{it}$ are zero, in which case $\mathbf{u}$ contains only martingales. In each case the definition of the process $\mathbf{u}$ is otherwise *model-free*. The underlying process $\mathbf{x}$ *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 $\mathbf{u}(T, \mathbf{x}_T) = \psi(\mathbf{x}_T)$ the solution to (PDE II) is given by

$$\mathbf{u}(t, \mathbf{x}_t) = \mathbb{E}_t \left[\psi(\mathbf{x}_T) + \frac{1}{2} \int_t^T \left(\text{tr} \left\{ \boldsymbol{\Sigma}_\tau^\top \boldsymbol{\Delta}_\tau^\top \mathbf{C}_{1\tau} \boldsymbol{\Delta}_\tau \boldsymbol{\Sigma}_\tau \right\}, \dots, \text{tr} \left\{ \boldsymbol{\Sigma}_\tau^\top \boldsymbol{\Delta}_\tau^\top \mathbf{C}_{n\tau} \boldsymbol{\Delta}_\tau \boldsymbol{\Sigma}_\tau \right\} \right)^\top d\tau \right].$$

Hence, if $\mathbf{C}_{it} = \mathbf{0}$, the derivative process u_i follows a martingale.

Proof of Corollary 1: Define

$$\mathbf{y}_s = \mathbf{u}(s, \mathbf{x}_s) + \frac{1}{2} \int_t^s \left(\text{tr} \left\{ \boldsymbol{\Sigma}_\tau^\top \boldsymbol{\Delta}_\tau^\top \mathbf{C}_{1\tau} \boldsymbol{\Delta}_\tau \boldsymbol{\Sigma}_\tau \right\}, \dots, \text{tr} \left\{ \boldsymbol{\Sigma}_\tau^\top \boldsymbol{\Delta}_\tau^\top \mathbf{C}_{n\tau} \boldsymbol{\Delta}_\tau \boldsymbol{\Sigma}_\tau \right\} \right)^\top d\tau,$$

so that $\mathbf{y}_t = \mathbf{u}(t, \mathbf{x}_t)$. By Itô's lemma,

$$d\mathbf{y}_s = d\mathbf{u}_s + \frac{1}{2} \left(\text{tr} \left\{ \boldsymbol{\Sigma}_s^\top \boldsymbol{\Delta}_s^\top \mathbf{C}_{1s} \boldsymbol{\Delta}_s \boldsymbol{\Sigma}_s \right\}, \dots, \text{tr} \left\{ \boldsymbol{\Sigma}_s^\top \boldsymbol{\Delta}_s^\top \mathbf{C}_{ns} \boldsymbol{\Delta}_s \boldsymbol{\Sigma}_s \right\} \right)^\top ds.$$

Substituting for $d\mathbf{u}_s$ using (4), and (8) with $\boldsymbol{\alpha}_s = \mathbf{0}$, $\forall s$, yields $d\mathbf{y}_s = \boldsymbol{\Delta}_s \boldsymbol{\Sigma}_s d\mathbf{w}_s$, i.e. $\mathbf{y}_s$ is a martingale. Hence

$$\mathbf{y}_t = \mathbb{E}_t[\mathbf{y}_T] = \mathbb{E}_t \left[\psi(\mathbf{x}_T) + \frac{1}{2} \int_t^T \left(\text{tr} \left\{ \boldsymbol{\Sigma}_\tau^\top \boldsymbol{\Delta}_\tau^\top \mathbf{C}_{1\tau} \boldsymbol{\Delta}_\tau \boldsymbol{\Sigma}_\tau \right\}, \dots, \text{tr} \left\{ \boldsymbol{\Sigma}_\tau^\top \boldsymbol{\Delta}_\tau^\top \mathbf{C}_{n\tau} \boldsymbol{\Delta}_\tau \boldsymbol{\Sigma}_\tau \right\} \right)^\top d\tau \right]$$

and on substituting the boundary condition $\mathbf{u}(T, \mathbf{x}_T) = \psi(\mathbf{x}_T)$ we have $\mathbf{y}_t = \mathbf{u}(t, \mathbf{x}_t)$. $\square$

Theorem 2: Suppose that $\mathbf{C}_{it} = \mathbf{Q} \mathbf{D}_{it} \mathbf{Q}^\top$, where $\mathbf{Q}$ is orthogonal and $\mathbf{D}_{it}$ are diagonal matrices, and that $\sum_{i=1}^n \mathbf{C}_{it} = (\mathbf{C}_{1t} \mathbf{1}, \dots, \mathbf{C}_{nt} \mathbf{1})^\top$. Then the solution to (PDE I) is given by

$$f(\mathbf{u}_r, \mathbf{u}_s) = a(\mathbf{u}_s) - a(\mathbf{u}_r) + \mathbf{b}(\mathbf{u}_r)^\top \{\mathbf{m}(\mathbf{u}_s) - \mathbf{m}(\mathbf{u}_r)\},$$

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

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

is a multivariate martingale.

Remark: Again, the a term satisfies the (AP) trivially, and the $\mathbf{b}$ term has zero expectation. If we do not impose $a \in \mathcal{C}^2$ and $\mathbf{b} \in \mathcal{C}^1$ there may be other functions f which satisfy the AP. Now the following Corollary characterises all solutions to (PDE I) and (PDE II) in the case that $\mathbf{u}$ follows a multivariate log martingale.

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

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

and $\mathbf{H}_i^m(\mathbf{u}_t) = \mathbf{C}_{it} \mathbf{J}^m(\mathbf{u}_t) = \mathbf{J}^m(\mathbf{u}_t) \mathbf{C}_{it}$. Now consider

$$\begin{aligned} & \left(\frac{\partial}{\partial u_{1s}}, \dots, \frac{\partial}{\partial u_{ns}} \right)^\top \left(\{\mathbf{J}(\mathbf{u}_r, \mathbf{u}_s) - \mathbf{J}^a(\mathbf{u}_s)\}^\top \mathbf{J}^m(\mathbf{u}_s)^{-1} \right) \\ &= \left(\mathbf{H}(\mathbf{u}_r, \mathbf{u}_s) - \mathbf{H}^a(\mathbf{u}_s) - \{\mathbf{J}(\mathbf{u}_r, \mathbf{u}_s) - \mathbf{J}^a(\mathbf{u}_s)\}^\top (\mathbf{C}_{1s}, \dots, \mathbf{C}_{ns}) \right) \mathbf{J}^m(\mathbf{u}_s)^{-1} \\ &= \left(\mathbf{H}(\mathbf{u}_r, \mathbf{u}_s) - \mathbf{H}^a(\mathbf{u}_s) - \sum_{i=1}^n \{J_i(\mathbf{u}_r, \mathbf{u}_s) - J_i^a(\mathbf{u}_s)\} \mathbf{C}_{is} \right) \mathbf{J}^m(\mathbf{u}_s)^{-1} = \mathbf{0}, \end{aligned}$$

where we apply the Lemma in the appendix in the second line and (PDE I) in the third line, and hence

$$\{\mathbf{J}(\mathbf{u}_r, \mathbf{u}_s) - \mathbf{J}^a(\mathbf{u}_s)\}^\top \mathbf{J}^m(\mathbf{u}_s)^{-1} = \mathbf{b}(\mathbf{u}_r)^\top,$$

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

$$f(\mathbf{u}_r, \mathbf{u}_s) = a(\mathbf{u}_s) - a(\mathbf{u}_r) + \mathbf{b}(\mathbf{u}_r)^\top \{\mathbf{m}(\mathbf{u}_s) - \mathbf{m}(\mathbf{u}_r)\},$$

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

$$d\mathbf{m}(\mathbf{u}_t) = \mathbf{J}^m(\mathbf{u}_t) d\mathbf{u}_t + \frac{1}{2} \mathbf{J}^m(\mathbf{u}_t) (\text{tr} \{\mathbf{C}_{1t} d\langle \mathbf{u} \rangle_t\}, \dots, \text{tr} \{\mathbf{C}_{nt} d\langle \mathbf{u} \rangle_t\})^\top = \mathbf{J}^m(\mathbf{u}_t) \boldsymbol{\Delta}_t \boldsymbol{\Sigma}_t d\mathbf{w}_t,$$

and therefore $\mathbf{m}$ is a multivariate martingale. $\square$

Corollary 2: Suppose that $\mathbf{C}_{it} = \mathbf{C}_i$ are constant and that $\sum_{i=1}^n \mathbf{C}_i$ is invertible. Then

$$\mathbf{m}(\mathbf{u}_t) = \left(\sum_{i=1}^n \mathbf{C}_i \right)^{-1} \left(\exp \left\{ \sum_{i=1}^n \mathbf{C}_i u_{it} \right\} - \mathbf{I} \right) \mathbf{1},$$

and

$$\mathbf{u}(t, \mathbf{x}_t) = \left(\sum_{i=1}^n \mathbf{C}_i \right)^{-1} \ln \left(\mathbb{E}_t \left[\exp \left\{ \sum_{i=1}^n \mathbf{C}_i \psi_i(\mathbf{x}_T) \right\} \right] \right) \mathbf{1},$$

the log of a martingale satisfying the boundary condition $\mathbf{u}_T = \boldsymbol{\psi}(\mathbf{x}_T)$.

Proof of Corollary 2: Note that

$$\mathbf{m}(\mathbf{u}_t) = \int_0^{\mathbf{u}_t} \exp \left\{ \sum_{i=1}^n \mathbf{C}_i u_i \right\} d\mathbf{u} = \left(\sum_{i=1}^n \mathbf{C}_i k_i \right)^{-1} \left(\exp \left\{ \sum_{i=1}^n \mathbf{C}_i u_{it} \right\} - \mathbf{I} \right) \mathbf{k}, \quad (14)$$

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

$$\mathbf{m}(\mathbf{u}_T) = \left(\sum_{i=1}^n \mathbf{C}_i \right)^{-1} \left(\exp \left\{ \sum_{i=1}^n \mathbf{C}_i \psi_i(\mathbf{x}_T) \right\} - \mathbf{I} \right) \mathbf{1}.$$

Now $\mathbf{m}$ is a martingale and we have $\mathbf{m}(\mathbf{u}_t) = \mathbb{E}_t[\mathbf{m}(\mathbf{u}_T)]$, i.e.

$$\mathbf{m}(\mathbf{u}_t) = \left(\sum_{i=1}^n \mathbf{C}_i k_i \right)^{-1} \mathbb{E}_t \left[\exp \left\{ \sum_{i=1}^n \mathbf{C}_i \psi_i(\mathbf{x}_T) \right\} - \mathbf{I} \right] \mathbf{1}. \quad (15)$$

Secondly, since $\sum_{i=1}^n \mathbf{C}_i k_i \mathbf{m} = \sum_{i=1}^n \mathbf{C}_i m_i \mathbf{k}$ by the Lemma in the appendix, we can write $\mathbf{u}$ as a function of $\mathbf{m}$ rather than the converse, using the equivalent expression to (14), i.e.

$$\mathbf{u}_t = \left(\sum_{i=1}^n \mathbf{C}_i \right)^{-1} \ln \left(\mathbf{I} + \sum_{i=1}^n \mathbf{C}_i m_{it} \right) \mathbf{1}, \quad (16)$$

where the choice of $\mathbf{1}$ is again arbitrary. Finally, inserting (15) in (16) and once again making use of the Lemma yields

$$\mathbf{u}(t, \mathbf{x}_t) = \left(\sum_{i=1}^n \mathbf{C}_i \right)^{-1} \ln \left(\mathbb{E}_t \left[\exp \left\{ \sum_{i=1}^n \mathbf{C}_i \psi_i(\mathbf{x}_T) \right\} \right] \right) \mathbf{1},$$

the log of a martingale satisfying the boundary condition $\mathbf{u}_T = \boldsymbol{\psi}(\mathbf{x}_T)$. $\square$

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

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

$$f(\mathbf{u}_r, \mathbf{u}_s) = a(\mathbf{u}_s) - a(\mathbf{u}_r) + \mathbf{b}(\mathbf{u}_r)^\top (\mathbf{u}_s - \mathbf{u}_r), \quad (\text{M})$$

where we assume w.l.o.g. that $\mathbf{J}^a(\mathbf{0}) = \mathbf{0}$. The solution to (PDE II) is $\mathbf{u}(t, \mathbf{x}_t) = \mathbb{E}_t[\boldsymbol{\psi}(\mathbf{x}_T)]$.

Proof of Corollary 3: We can either derive (M) as the limit for $\mathbf{C}_i \rightarrow \mathbf{0}$ 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

$$f(\mathbf{u}_r, \mathbf{u}_s) = a(\mathbf{u}_s) - a(\mathbf{u}_r) + \mathbf{b}(\mathbf{u}_r)^\top \left(\sum_{i=1}^n \mathbf{C}_i \right)^{-1} \sum_{i=1}^n \mathbf{C}_i (u_{is} - u_{ir}) \mathbf{1} + O \left(\left| \sum_{i=1}^n \mathbf{C}_i \right| \right),$$

where $\sum_{i=1}^n \mathbf{C}_i$ cancels out in the $\mathbf{b}$ term (after applying the Lemma in the appendix) and the higher orders vanish as $\mathbf{C}_i \rightarrow \mathbf{0}$, $i = 1, \dots, n$. For the latter derivation, note that setting all $\mathbf{C}_{is} = \mathbf{0}$ in Theorem 1 yields $\mathbf{H}(\mathbf{u}_r, \mathbf{u}_s) = \mathbf{H}^a(\mathbf{u}_s)$. Then

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

and integrating yields

$$\int_{\mathbf{u}_r}^{\mathbf{u}_s} [\mathbf{J}(\mathbf{u}_r, \mathbf{u}_t) - \mathbf{J}(\mathbf{u}_r, \mathbf{u}_r)]^\top d\mathbf{u}_t = \int_{\mathbf{u}_r}^{\mathbf{u}_s} [\mathbf{J}^a(\mathbf{u}_t) - \mathbf{J}^a(\mathbf{u}_r)]^\top d\mathbf{u}_t.$$

Integrating the above once again and using $f(\mathbf{u}, \mathbf{u}) = 0$ yields the solution

$$f(\mathbf{u}_r, \mathbf{u}_s) = a(\mathbf{u}_s) - a(\mathbf{u}_r) + \mathbf{b}(\mathbf{u}_r)^\top (\mathbf{u}_s - \mathbf{u}_r),$$

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

Clearly, the univariate martingale case corresponds to the solutions of (APB) found by Bondarenko [2014]. However, is not straightforward to show that the solution (M) encompasses the aggregating second and third moments found by Neuberger [2012]. This is given in the appendix (Corollary 4).

Next we expand upon the role of the functions a and $\mathbf{b}$ in defining the aggregating characteristic (M). First note that the implied characteristic is defined by a alone. This is because the realised characteristic based on (M), i.e.

$$\sum_{i=1}^N f(\mathbf{u}_{i-1}, \mathbf{u}_i) = a(\mathbf{u}_T) - a(\mathbf{u}_0) + \sum_{i=1}^N \mathbf{b}(\mathbf{u}_{i-1})^\top (\mathbf{u}_i - \mathbf{u}_{i-1}),$$

is an unbiased estimator for the implied characteristic $\mathbb{E}_0[f(\mathbf{u}_0, \mathbf{u}_T)]$ which, by virtue of $\mathbf{u}$ being a martingale, may also be written $\mathbb{E}_0[a(\mathbf{u}_T)] - a(\mathbf{u}_0)$.

While a must be fixed at time 0, $\mathbf{b}$ can change dynamically over time. Moreover, $\mathbf{b}$ determines the conditional variance of the estimator along the partition, because

$$\begin{aligned} & \mathbb{E}_t \left[\left\{ \sum_{i=1}^N f(\mathbf{u}_{i-1}, \mathbf{u}_i) - \mathbb{E}_t \left[\sum_{i=1}^N f(\mathbf{u}_{i-1}, \mathbf{u}_i) \right] \right\}^2 \right] \\ &= \mathbb{E}_t \left[\left\{ a(\mathbf{u}_T) - \mathbb{E}_t[a(\mathbf{u}_T)] + \sum_{i>t} \mathbf{b}(\mathbf{u}_{i-1})^\top (\mathbf{u}_i - \mathbf{u}_{i-1}) \right\}^2 \right], \end{aligned}$$

for $t \in \{t_0, \dots, t_N\}$. Next we propose an optimal choice for $\mathbf{b}$ 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 $\mathbb{E}_0[f(\mathbf{u}_0, \mathbf{u}_T)]$ has a first-order approximation given by $\mathbf{b}^*(\mathbf{u}) = -\mathbf{J}^a(\mathbf{u})$. If the process is a diffusion and monitoring happens continuously, the approximation is exact.

Proof of Theorem 3: For $t \in \{t_0, \dots, t_N\}$ we are interested in minimising the expression

$$\begin{aligned}
& \mathbb{E}_t \left[\left\{ \sum_{i=1}^N f(\mathbf{u}_{i-1}, \mathbf{u}_i) - \mathbb{E}_t \left[\sum_{i=1}^N f(\mathbf{u}_{i-1}, \mathbf{u}_i) \right] \right\}^2 \right] \\
&= \mathbb{E}_t \left[\left\{ a(\mathbf{u}_T) - \mathbb{E}_t[a(\mathbf{u}_T)] + \sum_{t_i > t} \mathbf{b}(\mathbf{u}_{i-1})^\top (\mathbf{u}_i - \mathbf{u}_{i-1}) \right\}^2 \right] \\
&= \mathbb{E}_t \left[a(\mathbf{u}_T)^2 \right] - \mathbb{E}_t[a(\mathbf{u}_T)]^2 + 2 \sum_{t_i > t} \mathbb{E}_t \left[a(\mathbf{u}_T) \mathbf{b}(\mathbf{u}_{i-1})^\top (\mathbf{u}_i - \mathbf{u}_{i-1}) \right] \\
&\quad + \sum_{t_i > t} \mathbb{E}_t \left[\left\{ \mathbf{b}(\mathbf{u}_{i-1})^\top (\mathbf{u}_i - \mathbf{u}_{i-1}) \right\}^2 \right] \\
&= \mathbb{E}_t \left[a(\mathbf{u}_T)^2 \right] - \mathbb{E}_t[a(\mathbf{u}_T)]^2 + 2 \sum_{t_i > t} \mathbb{E}_t \left[\mathbf{b}(\mathbf{u}_{i-1})^\top \mathbb{E}_{t_{i-1}}[a(\mathbf{u}_T) (\mathbf{u}_i - \mathbf{u}_{i-1})] \right] \\
&\quad + \sum_{t_i > t} \mathbb{E}_t \left[\mathbf{b}(\mathbf{u}_{i-1})^\top \mathbb{E}_{t_{i-1}} \left[(\mathbf{u}_i - \mathbf{u}_{i-1}) (\mathbf{u}_i - \mathbf{u}_{i-1})^\top \right] \mathbf{b}(\mathbf{u}_{i-1}) \right] \\
&= \mathbb{E}_t \left[a(\mathbf{u}_T)^2 \right] - \mathbb{E}_t[a(\mathbf{u}_T)]^2 + \sum_{t_i > t} \mathbb{E}_t \left[2\mathbf{b}(\mathbf{u}_{i-1})^\top \boldsymbol{\omega}_{i-1} + \mathbf{b}(\mathbf{u}_{i-1})^\top \boldsymbol{\Omega}_{i-1} \mathbf{b}(\mathbf{u}_{i-1}) \right],
\end{aligned}$$

with $\boldsymbol{\omega}_{i-1} = \mathbb{E}_{t_{i-1}}[a(\mathbf{u}_T) (\mathbf{u}_i - \mathbf{u}_{i-1})]$ and $\boldsymbol{\Omega}_{i-1} = \mathbb{E}_{t_{i-1}}[(\mathbf{u}_i - \mathbf{u}_{i-1}) (\mathbf{u}_i - \mathbf{u}_{i-1})^\top]$. Taking the derivative with respect to the components of $\mathbf{b}(\mathbf{u}_t)$ and setting the result equal to zero yields $\mathbf{b}(\mathbf{u}_t) = -\boldsymbol{\Omega}_t^{-1} \boldsymbol{\omega}_t$. The second derivative w.r.t. $\mathbf{b}$ corresponds to $\boldsymbol{\Omega}$ and is positive definite as long as the components of $\boldsymbol{\psi}$ are linearly independent. We have therefore found a unique minimum. Assume now that $\mathbf{u}$ follows the dynamics $d\mathbf{u}_t = \boldsymbol{\Delta}_t \boldsymbol{\Sigma}_t d\mathbf{w}_t$ and, for some $r < s$, consider $\boldsymbol{\omega}_r = \mathbb{E}_r[(\mathbb{E}_t[a(\mathbf{u}_T)] - \mathbb{E}_r[a(\mathbf{u}_T)])(\mathbf{u}_s - \mathbf{u}_r)]$ and $\boldsymbol{\Omega}_r = \mathbb{E}_r[(\mathbf{u}_s - \mathbf{u}_r)(\mathbf{u}_s - \mathbf{u}_r)^\top]$. Then, as $r \rightarrow s$ (continuous monitoring), we have $\boldsymbol{\Omega}_t = d\langle \mathbf{u} \rangle_t$ and $\boldsymbol{\omega}_t = d\langle \mathbf{u} \rangle_t \mathbf{J}^a(\mathbf{u}_t)$ (using that $d\mathbb{E}_t[a(\mathbf{u}_T)] = \mathbf{J}^a(\mathbf{u}_t)^\top d\mathbf{u}_t$), and therefore $\mathbf{b}(\mathbf{u}) = -\boldsymbol{\Omega}^{-1} \boldsymbol{\omega} = -\mathbf{J}^a(\mathbf{u})$.

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

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 (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.

In the next section we consider some particular choices for a and $\mathbf{b}$ which correspond to unbiased and efficient higher-moment estimators. Then we provide a numerical comparison of the new estimators with standard higher-moment estimators as well as the second and third moment estimators by Neuberger. Furthermore, we use a 24 years history of S&P 500 futures and options data to illustrate the historical performance of these higher moments.

3 Aggregating Higher Moments

Suppose we restrict $\mathbf{u}$ to contain a series of power log contracts $P_t^{(i)} = \mathbb{E}_t [y_T^i]$, $i \geq 0$ on a single underlying F , where $y_t = \ln F_t$ for $0 \leq t \leq T$. Applying the replication theorem of Carr and Madan [2001], for $i \geq 1$ this conditional expectation can be expressed as:

$$P_t^{(i)} = y_t^i + \int_{\mathbb{R}^+} \gamma_i(k) q_t(k) dk, \quad (17)$$

where $q_t(k)$ denotes the time- t price of a vanilla out-of-the-money (OTM) option with strike k and maturity T and $\gamma_i(k) := i(\ln k)^{i-2} k^{-2} [i - 1 - \ln k]$. For instance, when $i = 1$ we obtain the standard replication portfolio $Y_t = y_t - \int_{\mathbb{R}^+} k^{-2} q_t(k) dk$ for the log contract $Y_t = \mathbb{E}_t [y_T]$. Note that $P^{(0)} = 1$, and that we write $P^{(1)} = Y$ for convenience of notation.

Now let $\mathbf{u} = (1, Y, P^{(2)} \dots, P^{(n-1)})^\top$ for $n \geq 2$ and consider the specification

$$a(\mathbf{u}) = - \sum_{i=0}^{n-1} \binom{n}{i} P^{(i)} (-Y)^{n-i}. \quad (18)$$

Then $a(\mathbf{u}_T) = y_T^n$, since $P_T^{(i)} = y_T^i$ and $\sum_{i=0}^n \binom{n}{i} (-1)^{n-i} = 0$, and according to Corollary 3 the implied characteristic equals

$$\mathbb{E}_0 [f(\mathbf{u}_0, \mathbf{u}_T)] = \mathbb{E}_0 [a(\mathbf{u}_T)] - a(\mathbf{u}_0) = \sum_{i=0}^n \binom{n}{i} P_0^{(i)} (-Y_0)^{n-i} = \mathbb{E}_0 [(y_T - Y_0)^n] \quad (19)$$

which is the n^{th} central moment of the log return distribution.

For $n = 2$ we have $a(\mathbf{u}) = Y^2$ and, using Theorem 3, $\mathbf{b}^*(\mathbf{u}) = -2Y$. Hence, our realised second moment (RSM) is the squared change in price of the log contract:⁸

$$f(\mathbf{u}_r, \mathbf{u}_s) = (Y_s - Y_r)^2. \quad (\text{RSM})$$

It is replicated by holding a squared log contract and shorting $2Y_r$ log contracts between times r and s . The calculation of Y_t from option prices has a numerical integration error, whereas Neuberger's

⁸This corresponds to the c_8 term in Neuberger's geometric set of solutions described in the appendix.

log variance (LV) may be derived from direct observation of the underlying price alone. However, there is very little difference between them, as shown in Figure 1. This figure compares (RSM) with existing definitions of RV, 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. Also note that here, and in the following, we set $s - r = 1/250$ for daily monitoring.

Figure 1: Comparison of our efficient realised second moment (RSM), Neuberger's log variance (LV) as well as the conventional squared log return (CONV). The left diagram shows the three estimators for $s - r = 1/250$ (daily monitoring) and $\sigma = 20\%$ (constant implied volatility), where the x -axis corresponds to $y_s - y_r$, i.e. the log return. Note that in this setting we have $y_s - y_r = Y_s - Y_r + \frac{\sigma^2}{2}(s - r)$. The right diagram shows the deviations from the conventional estimator, where bp = 10^{-4} stands for basis points.

Now suppose $n \geq 3$ and again apply Theorem 3 to derive the first-order efficient estimator for the n^{th} central moment of the log return distribution as:

$$\mathbf{b}^*(\mathbf{u}) = \left(n(-Y)^{n-1} - \sum_{i=0}^{n-1} \binom{n}{i} (n-i) P^{(i)} (-Y)^{n-1-i}, \dots, n(-Y) \right)^\top. \quad (20)$$

Setting $n = 3$ in (18) and (20) yields

$$a(\mathbf{u}) = -2Y^3 + 3P^{(2)}Y \quad \text{and} \quad \mathbf{b}^*(\mathbf{u}) = \left(6Y^2 - 3P^{(2)}, -3Y \right)^\top.$$

Therefore, if we define the n^{th} central moment process $v_t^{(n)} = \mathbb{E}_t[(y_T - \mathbb{E}_t[y_T])^n]$, and recalling our notation $Y_t = \mathbb{E}_t[y_T]$ we can express our new realised third moment (RTM) as:⁹

$$f(\mathbf{u}_r, \mathbf{u}_s) = (Y_s - Y_r)^3 + 3 \left(v_s^{(2)} - v_r^{(2)} \right) (Y_s - Y_r). \quad (\text{RTM})$$

⁹Note that this is very similar to the c_4 term in Neuberger's arithmetic set of solutions, but it 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 .

By (19), RTM is an aggregating third moment for which the implied characteristic corresponds *exactly* to the third central moment of the log return distribution – unlike Neuberger’s third moment (NTM) which contains an additional price-variance correlation term. The realised third moment (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.

In Figure 2 we compare the two realised third moment measures with the conventional third moment, i.e. $(y_s - y_r)^3$, the cubed log return (CONV). As expected, the RTM is closer to CONV than NTM. Also RTM is marginally below CONV 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.

Figure 2: Comparison of our efficient realised third moment (RTM), Neuberger’s third moment (NTM), as well as the cubed log return (CONV), i.e. $(y_s - y_r)^3$. The left diagram shows the three estimators for $s - t = 1/250$ (daily monitoring) and $\sigma = 20\%$ (constant implied volatility), where again the x -axis corresponds to $y_s - y_r = Y_s - Y_r + \frac{\sigma^2}{2}(s - r)$ and $\text{bp} = 10^{-4}$ are basis points. Note that $v^\eta \approx v^{(2)} + v^{(3)}/3$, 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 $v^{(3)} = 0$ for simplicity. The right diagram shows the deviations from the conventional estimator.

Finally we set $n = 4$ in (18) and (20) to obtain a fourth-moment characteristic with $a(\mathbf{u}) = 3Y^4 - 6P^{(2)}Y^2 + 4P^{(3)}Y$ and $\mathbf{b}^*(\mathbf{u}) = (-12Y^3 + 12P^{(2)}Y - 4P^{(3)}, 6Y^2, -4Y)^\top$ so:

$$f(\mathbf{u}_r, \mathbf{u}_s) = (Y_s - Y_r)^4 + 6v_s^{(2)}(Y_s - Y_r)^2 + 4(v_s^{(3)} - v_r^{(3)})(Y_s - Y_r). \quad (\text{RFM})$$

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 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 (19), 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 moment, 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 $|Y_s - Y_r|$.

Figure 4 extends this numerical comparison between conventional and time-aggregating higher moments to empirical data. It shows the actual realisations of higher moments derived from 24-years of S&P 500 stock market data – more specifically, daily closing futures and options prices from April 1995 to March 2019. Standard no-arbitrage interpolation techniques are applied to synthesise power log contracts from these data, and all contracts have constant 30-days to maturity by construction. The value of any dot on the x-axis corresponds to $y_s - y_r$, the log return on a given day, while the value on the y-axis corresponds to the respective realised higher moment formula, depending on their color. In the left-hand figure, realisations of the log variance (LV) and the squared log return (CONV) are just deterministic functions of the log return. However, the RSM shown by the red dots is stochastic because (by contrast with the numerical examples above) the implied volatility is stochastic, in real financial data. Nevertheless, the correlation between RSM and the

Figure 3: Comparison of our efficient RFM with the fourth power of the log return (CONV), i.e. $(y_s - y_r)^4$. The left diagram shows the two estimators for $s - r = 1/250$ (daily monitoring) and $\sigma = 20\%$ (constant implied volatility), where again the x -axis corresponds to $y_s - y_r = Y_s - Y_r + \frac{\sigma^2}{2}(s - r)$ and $\text{bp} = 10^{-4}$ are basis points. 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. The right diagram shows the deviation from the conventional estimator.

other two measures is well over 99%. The right-hand figure reveals that the fourth power of the log return (CONV) only captures a small proportion of the features of a time-aggregating fourth-moment estimator. In fact, our realised fourth moment (RFM) is mostly the result of volatility clustering, with little evidence of additional, asymmetric leverage effects. Nevertheless, the empirical correlation between RFM and the conventional measure is 81.39%. However, it is the centre chart where the impact of changes in the stochastic implied volatility on higher-order realised moments matters most. In particular, the cubed log return (CONV) completely fails to capture most of the features of the two time-aggregating third moments RTM and NTM. Whilst these two moment estimators yield estimates with correlation above 99%, the correlations of RTM and NTM with the conventional third moment estimates are -15.5% and -22.7%, respectively.

Figure 4: Daily realised higher moments of the S&P 500 between April 1995 and March 2019, based on both conventional and aggregating measures.

4 Frequency Swaps

A frequency swap is a financial instrument that exchanges a pay-off, monitored at two different frequencies, whose fair value is independent of the monitoring frequency. For example, when realised variance (RV) is measured using our aggregating second-moment characteristic, the risk-neutral expectation (i.e. the fair value) is independent of the frequency over which RV is monitored. Nevertheless, in the physical measure, RV does depend on the monitoring frequency – typically, when the RV is monitored daily over some time period (i.e. the maturity of the swap) it will be greater than when it is monitored at the weekly frequency, over the same maturity. Therefore, the pay-off to a variance swap struck at the fair-value swap rate (i.e. the VRP) will also depend on the monitoring frequency.

Figure 5 illustrates this using 24 years of daily closing prices on S&P 500 futures with a constant maturity of 30 days. We also use the corresponding closing prices of standard European options to derive the fair-value swap rate, which is independent of the monitoring frequency of the realised leg of the swap. The upper graph depicts the VRP when RV is monitored daily, weekly and monthly.¹⁰ The dependence on monitoring frequency is clear and the (typically) negative VRP only becomes positive during excessively volatile periods such as the banking crisis of 2008-9.

Although frequency swaps are not traded in high volumes, variance swaps are one of the most highly traded over-the-counter instruments in equity (and other) markets. Now, frequency swaps occur naturally as the residual hedging exposure for traders of an aggregating variance swap (or indeed any swap contract which is defined by an aggregating pay-off) when hedging is at a frequency that differs from the monitoring of the pay-off. Of course, traders could hedge their exposure perfectly by trading in the vector $\mathbf{u}$ of derivative assets every time the swap is monitored but, given transaction costs, it is often more practical for hedging to be less frequent than the monitoring frequency. For example, it may be convenient to sell a daily monitored swap but only hedge it once a week. To see this, suppose the monitoring partition is π_m but hedging is done at the partition

¹⁰By convention, a week consists of 5 trading days and a month consists of 20 trading days. Each time after monitoring has taken place, the maturity of all contracts is reset to 30 days. Note that, in the case of monthly monitoring, there are always 20 trading days within the 30-days constant maturity window.

Figure 5: The upper graph displays the cumulative premium earned on a variance swap on the S&P 500 with daily, weekly and monthly monitoring over a 24-years period. The second graph displays the cumulative differences, corresponding to a weekly-for-monthly and a daily-for-weekly frequency swap, respectively. Clearly, the term structure of realised variance is decreasing in the sense that RV measured over a wider partition is lower than RV measured over a finer partition of the same time interval.

π_h , where $\pi_m \supset \pi_h$. The residual exposure corresponds to a frequency swap with the floating leg

$$\sum_{\pi_m} f(\mathbf{u}_{t_{i-1}}, \mathbf{u}_{t_i}) - \sum_{\pi_h} f(\mathbf{u}_{t_{i-1}}, \mathbf{u}_{t_i}).$$

The AP implies

$$\mathbb{E}_{t_0} \left[\sum_{\pi_m} f(\mathbf{u}_{t_{i-1}}, \mathbf{u}_{t_i}) \right] = \mathbb{E}_{t_0} \left[\sum_{\pi_h} f(\mathbf{u}_{t_{i-1}}, \mathbf{u}_{t_i}) \right]$$

and, because the corresponding swap rates for the two floating components cancel out, the fair-value swap rate of this frequency swap is zero at inception. However, for $t > t_0$ the P&L need not be zero in the presence of a hedging error. In fact, for $t \in \pi_h$ the mark-to-market P&L on this frequency swap is

$$\mathbb{E}_t \left[\sum_{\pi_m} f(\mathbf{u}_{t_{i-1}}, \mathbf{u}_{t_i}) - \sum_{\pi_h} f(\mathbf{u}_{t_{i-1}}, \mathbf{u}_{t_i}) \right] = \sum_{\pi_m \cap [t_0, t]} f(\mathbf{u}_{t_{i-1}}, \mathbf{u}_{t_i}) - \sum_{\pi_h \cap [t_0, t]} f(\mathbf{u}_{t_{i-1}}, \mathbf{u}_{t_i}).$$

In our empirical example of an aggregating variance swap, this residual premium is illustrated by

the lower graph in Figure 5 for a weekly-monitored swap which is hedged monthly (blue line), and for the daily-monitored swap which is hedged weekly (red line).

5 Conclusions

We characterise all solutions to a general aggregation property for expectations of non-linear functions f of a general multivariate stochastic processes $\mathbf{x}_t$. Our property encompasses, as different special cases, the specific bivariate case in Neuberger [2012] and the univariate martingale case of Bondarenko [2014]. Our solutions require only minimal assumptions on first and second derivatives of f and a derivative process $\mathbf{u}_t$ related to $\mathbf{x}_t$. In the special cases that $\mathbf{u}_t$ contains only martingales, or log martingales, we derive closed-form solutions that are fully determined by a real-valued function a and a vector-valued function $\mathbf{b}$. The flexibility to select these functions generates an entire vector space of solutions which includes an infinite number of non-linear functions which satisfy the aggregation property.

We apply our solutions to derive unbiased and efficient estimates and forecasts for higher-moment risk premia, explaining how to select a using power log contracts in $\mathbf{u}_t$ so that the expectations correspond exactly to n^{th} -order moments of log returns. By satisfying the aggregation property our risk premia estimates are unbiased and model free – unlike the variance risk premium derived from standard variance swaps – and they have no discretisation error. They are also independent of the sampling frequency, i.e. these realised moments are the same whether derived from daily, weekly, monthly or even annual returns. We also show how to select $\mathbf{b}$ to derive efficient estimators which, unlike the third moment found by Neuberger [2012], correspond exactly to the n^{th} central moment of log returns.

A numerical comparison between our time-aggregating higher moments and other estimators is augmented by some applications based on real time-series data, i.e. 24 years of daily data on S&P 500 futures and options. First, corroborating Kozhan et al. [2013], we demonstrate that the conventional measures of short-term third moments based on high-frequency data cannot be used to infer anything about the properties of long-term third moments based on low-frequency data. Our results provide two new insights to this problem: by deriving a more canonical aggregating third moment and by introducing a new aggregating fourth moment, thus extending this finding about the inadequacy of the conventional third moment to the fourth moment.

The implications of this paper apply to managers of portfolios of financial assets and instruments. First consider the long-term asset managers, which typically have large institutional investors as clients. Such managers will re-optimize portfolios at quite low frequencies (typically monthly, quarterly or even annually) yet their optimized weights are derived from short-term data (typically daily or weekly). However, we show that nothing can be inferred about the long-term risk premia when optimization is based on conventional measures of higher moments derived from high-frequency data. Instead, long-term asset managers should employ the aggregating measures that we define in this paper.

Finally, we explain how frequency swaps arise naturally as the residual hedging exposure to managers of shorter-term higher-moment risk premia and we examine the empirical properties of such frequency swaps when hedging is performed at a lower frequency than monitoring of the floating leg of the swap. This has broad implications for short-term portfolio managers, such as hedge funds, who will typically hedge at a lower frequency than the monitoring frequency of the financial instruments that they trade.

Acknowledgements

We would like to thank the participants at EURO2018 in Valencia and seminar attendees at the Fields Institute in Toronto, the Courant Institute of Mathematical Sciences in New York, the University of Oxford Mathematics Institute, the Bachelier Conference in New York, Imperial College, University College London, Instituto de Matemática Pura e Aplicada (IMPA) in Rio de Janeiro and several practitioner conferences. The following individuals made very helpful comments on earlier drafts: Peter Carr, Michael Coulon, Bruno Dupire, Andreas Kaeck, Tobias Kuna, Markus Leippold, Matt Lorig, Anthony Neuberger, Yuri Saporito, Enrico Scalas, Matthias Scherer, Jorge Zubelli. All remaining errors are ours.

References

- H. Aghababa and W.A. Barnett. Dynamic structure of the spot price of crude oil: does time aggregation matter? *Energy Economics*, 59:227–237, 2016.
- Y. Ait-Sahalia. Disentangling diffusion from jumps. *Journal of Financial Economics*, 74:487–528, 2004.
- Y. Ait-Sahalia, M. Karaman, and L. Mancini. The term structure of variance swaps and risk premia. *Swiss Finance Institute Research Paper*, (18-37), 2018.
- G. Athanopoulos, R.J. Hyndman, N. Kourentzes, and F. Petropoulos. Forecasting with temporal hierarchies. *European Journal of Operational Research*, 262(1):60–74, 2017.
- S. Bastola and V. Misra. Sensitivity of hydrological simulations of southeastern united states watersheds to temporal aggregation of rainfall. *Journal of Hydrometeorology*, 14(4):1334–1344, 2013.
- 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.
- 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.
- C. Foroni, P. Guérin, and M. Marcellino. Using low frequency information for predicting high frequency variables. *International Journal of Forecasting*, 34(4):774–787, 2018.
- A. Forsman, C. Andersson, A. Grimvall, and M. Hoffmann. Estimation of the impact of short-term fluctuations in inputs on temporally aggregated outputs of process-oriented models. *Journal of Hydroinformatics*, 5(3):169–180, 2003.
- C.W.J. Granger. Non-linear models: Where do we go next - time varying parameter models? *Studies in Nonlinear Dynamics and Econometrics*, 12(3), 2008.
- C.M. Hafner. Causality and forecasting in temporally aggregated multivariate garch processes. *Econometrics Journal*, 12(1):127–146, 2009.
- 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.J. Hyndman, R.A. Ahmed, G. Athanasopoulos, and H.L. Shang. Optimal combination forecasts for hierarchical time series. *Computational Statistics and Data Analysis*, 55(9):2579–2589, 2011.
- 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.
- Y. Jin, B.D. Williams, T. Tokar, and M.A. Waller. Forecasting with temporally aggregated demand signals in a retail supply chain. *Journal of Business Logistics*, 36(2):199–211, 2015.
- E. Kole, T. Markwat, A. Opschoor, and D. van Dijk. Forecasting value-at-risk under temporal and portfolio aggregation. *Journal of Financial Econometrics*, 15(4):649–677, 2017.
- N. Kourentzes and F. Petropoulos. Forecasting with multivariate temporal aggregation: The case of promotional modelling. *International Journal of Production Economics*, 181:145–153, 2016.
- N. Kourentzes, F. Petropoulos, and J.R. Trapero. Improving forecasting by estimating time series structural components across multiple frequencies. *International Journal of Forecasting*, 30(2):291–302, 2014.
- 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.
- S. Leontsinis and C. Alexander. Arithmetic variance swaps. *Quantitative Finance*, 17(4):551–569, 2017.

- I.J. Lorite, L. Mateos, and E. Fereres. Impact of spatial and temporal aggregation of input parameters on the assessment of irrigation scheme performance. *Journal of Hydrology*, 300(1-4):286–299, 2005.
- N. Mamingi. Beauty and ugliness of aggregation over time: A survey. *Review of Economics*, 68(3):205–227, 2017.
- A. Neuberger. Realized skewness. *Review of Financial Studies*, 25(11):3423–3455, 2012.
- F. Petropoulos and N. Kourentzes. Forecast combinations for intermittent demand. *Journal of the Operational Research Society*, 66(6):914–924, 2015.
- 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.
- B. Rostami-Tabar, M.Z. Babai, M. Ali, and J.E. Boylan. The impact of temporal aggregation on supply chains with arma(1,1) demand processes. *European Journal of Operational Research*, 273(3):920–932, 2019.
- P. Schneider and F. Trojani. Divergence and the price of uncertainty. *Journal of Financial Econometrics*, 17(3):341–396, 2019.
- A. Silvestrini and D. Veredas. Temporal aggregation of univariate and multivariate time series models: A survey. *Journal of Economic Surveys*, 22(3):458–497, 2008.
- N. Taylor. The predictive value of temporally disaggregated volatility: Evidence from index futures markets. *Journal of Forecasting*, 27(8):721–742, 2008.
- H. Tsai and K.S. Chan. Temporal aggregation of stationary and nonstationary discrete-time processes. *Journal of Time Series Analysis*, 26(4):613–624, 2005.

Appendix

The proofs of Theorem 2 and its Corollaries require the following:

Lemma: Let $\mathbf{C}_i = \mathbf{Q}\mathbf{D}_i\mathbf{Q}^\top$ and $\sum_{i=1}^n \mathbf{C}_i = (\mathbf{C}_1\mathbf{1}, \dots, \mathbf{C}_n\mathbf{1})$. Then also

$$\sum_{i=1}^n \mathbf{C}_i \mathbf{k}_i = (\mathbf{C}_1\mathbf{k}, \dots, \mathbf{C}_n\mathbf{k}) = \left(\mathbf{C}_1^\top \mathbf{k}, \dots, \mathbf{C}_n^\top \mathbf{k} \right)^\top,$$

for all $\mathbf{k} = (k_1, \dots, k_n)^\top \in \mathbf{R}^n$. That is, the third order tensor $(\mathbf{C}_1, \dots, \mathbf{C}_n)$ is symmetric.

Proof of Lemma: Let $\mathbf{Q} = (\mathbf{q}_1, \dots, \mathbf{q}_n)^\top$ and $\mathbf{D}_i = \text{diag}(\mathbf{D}\mathbf{q}_i)$. Clearly $\mathbf{D}$ exists, and it is unique if $\mathbf{C}_i$ are linearly independent. Combining this with $\sum_{i=1}^n \mathbf{C}_i = (\mathbf{C}_1\mathbf{1}, \dots, \mathbf{C}_n\mathbf{1})$ yields

$$\sum_{i=1}^n \mathbf{Q} \text{diag}(\mathbf{D}\mathbf{q}_i) \mathbf{Q}^\top = \left(\mathbf{Q} \text{diag}(\mathbf{D}\mathbf{q}_1) \mathbf{Q}^\top \mathbf{1}, \dots, \mathbf{Q} \text{diag}(\mathbf{D}\mathbf{q}_n) \mathbf{Q}^\top \mathbf{1} \right).$$

Since $\text{diag}(\mathbf{D}\mathbf{q}_i) \mathbf{Q}^\top \mathbf{1} = \text{diag}(\mathbf{Q}^\top \mathbf{1}) \mathbf{D}\mathbf{q}_i$ we have

$$\mathbf{D} = \sum_{i=1}^n \text{diag}(\mathbf{D}\mathbf{q}_i) \text{diag}(\mathbf{Q}^\top \mathbf{1})^{-1} = \text{diag}(\mathbf{D}\mathbf{1}).$$

Therefore $\mathbf{D}$ must be diagonal. Then

$$\begin{aligned}
\sum_{i=1}^n \mathbf{C}_i k_i &= \sum_{i=1}^n \mathbf{Q} \operatorname{diag}(\mathbf{D}\mathbf{q}_i) \mathbf{Q}_i^\top k_i = \mathbf{Q} \operatorname{diag}(\mathbf{D}\mathbf{Q}^\top \mathbf{k}) (\operatorname{diag}\{\mathbf{q}_1\} \mathbf{1}, \dots, \operatorname{diag}\{\mathbf{q}_n\} \mathbf{1}) \\
&= (\mathbf{Q} \operatorname{diag}(\mathbf{D}\mathbf{q}_1) \mathbf{Q}^\top \mathbf{k}, \dots, \mathbf{Q} \operatorname{diag}(\mathbf{D}\mathbf{q}_n) \mathbf{Q}^\top \mathbf{k}) = (\mathbf{C}_1 \mathbf{k}, \dots, \mathbf{C}_n \mathbf{k}),
\end{aligned}$$

where we have used that $\operatorname{diag}(\mathbf{D}\mathbf{Q}^\top \mathbf{k}) \operatorname{diag}(\mathbf{q}_i) = \operatorname{diag}(\mathbf{D}\mathbf{q}_i) \operatorname{diag}\{\mathbf{Q}^\top \mathbf{k}\}$. Finally, since $\mathbf{C}_i$ are symmetric, we have $(\mathbf{C}_1 \mathbf{k}, \dots, \mathbf{C}_n \mathbf{k}) = (\mathbf{C}_1^\top \mathbf{k}, \dots, \mathbf{C}_n^\top \mathbf{k})^\top$. $\square$

Neuberger's Solutions: When $\mathbf{x} = F$, the forward price of a single tradeable asset, the only functions which satisfy (APN) are $g(\delta F) = \delta F$ and $g(\delta F) = (\delta F)^2$. Neuberger [2012] considers two bivariate cases: the 'arithmetic' case, $\mathbf{x} = (F, v)^\top$, where v_t is the conditional variance of F_T ; and the 'geometric' case, $\mathbf{x} = (y, v^\phi)^\top$, where $y = \ln F$ denotes the log forward price and v_t^ϕ is some 'generalised' variance process. This way he finds new second moments which, unlike the sum of squared log returns, satisfy (APN). He also finds one third moment for which (APN) holds,¹¹ and uses this to infer the skewness of long-term return distributions from observations on daily log returns.¹²

Neuberger [2012] characterises the solutions $(g; \mathbf{x})$ to (APN) only in the special case that $\mathbf{x}$ is bivariate and the second component is a conditional expectation of the first component, whence the process has an implicit dependence structure. First he considers the arithmetic case $\mathbf{x} = (F, v)^\top$, where $v_t = \mathbb{E}_t[(F_T - F_t)^2]$, deriving the solutions

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

where $\delta \mathbf{x}$ denotes an increment in $\mathbf{x}$ and $c_i \in \mathbb{R}$, $i \in \{1, \dots, 4\}$ are arbitrary real coefficients. The (c_1, c_2) term satisfies (APN) trivially, the c_3 term occurs provided F is a martingale (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 $\mathbf{x} = (y, v^\phi)^\top$, where $v_t^\phi = \mathbb{E}_t[\phi(y_T - y_t)]$ denotes a generalised variance process, i.e. any process such that $\lim_{\delta y \rightarrow 0} \phi(\delta y) / (\delta y)^2 = 1$. In this case he shows that the solutions to (APN) are given by

¹¹See equation (5) p.3430 for the arithmetic case and Proposition 6, p.3435 for the geometric case.

¹²Kozhan 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.

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

where $c_i \in \mathbb{R}$, $i \in \{5, \dots, 9\}$, 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^ϕ can be any generalised variance process. Within the geometric set of solutions to (APN), Neuberger [2012] focuses on the LV, for which $(c_5, c_6, c_7, c_8, c_9) = (-2, 0, 2, 0, 0)$, so that:¹³

$$g(\delta y) = 2(e^{\delta y} - 1 - \delta y) = \lambda(\delta y). \quad (21)$$

He finds only one higher-order moment, for $(c_5, c_6, c_7, c_8, c_9) = (6, -3, -12, 0, 3)$, given by

$$g(\delta \mathbf{x}) = \rho(\delta \mathbf{x}) + \tau(\delta y), \quad (22)$$

where $\rho(\delta \mathbf{x}) = 3\delta v^\eta(e^{\delta y} - 1)$ and $\tau(\delta y) = 6(\delta y e^{\delta y} - 2e^{\delta y} + \delta y + 2)$. We refer to (22) as Neuberger’s third moment (NTM). To find the implied moment corresponding to NTM, set $\delta \mathbf{x} = \mathbf{x}_T - \mathbf{x}_0$ and take expectations. Since $\mathbb{E}_0[\rho(\mathbf{x}_T - \mathbf{x}_0)] = 0$ when F is a martingale, we have $\mathbb{E}_0[g(\mathbf{x}_T - \mathbf{x}_0)] = \mathbb{E}_0[\tau(y_T - y_0)]$. Unfortunately, even though $\lim_{\delta y \rightarrow 0} \tau(\delta y) / (\delta y)^3 = 1$, the implied characteristic does not capture a third moment because it will be dominated by terms in τ with order greater than 3 when $y_T - y_0$ is sufficiently large. This motivates our search for a new third-moment solution to (AP) of order exactly 3.¹⁴

Finally, we show that the solution (M) encompasses NSM and NTM, the aggregating second and third moments found by Neuberger [2012].¹⁵ To this end we first re-write Neuberger’s geometric set of solutions in an unrestricted fashion by setting $\mathbf{x} = (y, v^\lambda, v^\eta)^\top$ and writing

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

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 $v^\eta = 2(\frac{Z}{F} - y)$.

¹³Note that $2(\delta y e^{\delta y} - e^{\delta y} + 1)$, i.e. the RV corresponding to the entropy variance process, does not satisfy (APN). That is, there is no choice of $(c_5, c_6, c_7, c_8, c_9)$ which yields $g(\delta y) = \eta(\delta y)$.

¹⁴The fact that $\tau(\delta y)$ is $O(\delta y^3)$ is useful for measurement based on high-frequency data. But $\rho(\delta \mathbf{x})$ is $O(\delta v^\eta \delta y)$, so $g(\delta \mathbf{x})$ is not a pure cubic exposure because it includes an additional price-variance correlation exposure. See Neuberger [2012] for further discussion on this point.

¹⁵Expressing 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.

Corollary 4: Let $\mathbf{u} = (F, Y, Z)^\top$ and set $a(\mathbf{u}) = (c_5 + 2c_6^\lambda - 2c_6^\eta, 4c_8, 2c_6^\eta) (\ln F, Y^2, \frac{Z}{F})^\top$ and $\mathbf{b}(\mathbf{u}) = (\frac{c_7}{F} - \frac{2c_9Z}{F^2}, -2c_6^\lambda - 8c_8Y, \frac{2c_9}{F})^\top$. Then

$$f(\mathbf{u}_r, \mathbf{u}_s) = a(\mathbf{u}_s) - a(\mathbf{u}_r) + \mathbf{b}(\mathbf{u}_r)^\top (\mathbf{u}_s - \mathbf{u}_r) = g(\mathbf{x}_s - \mathbf{x}_r),$$

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 $a(\mathbf{u}) = -2 \ln F$ and $\mathbf{b}(\mathbf{u}) = (\frac{2}{F}, 0, 0)^\top$, we have $f(\mathbf{u}_r, \mathbf{u}_s) = \lambda(y_s - y_r)$, which corresponds to one of Neuberger's aggregating second moments, the LV. Similarly, setting $(c_5, c_6^\lambda, c_6^\eta, c_7, c_8, c_9) = (6, 0, -3, -12, 0, 3)$ yields $a(\mathbf{u}) = 12 \ln F - 6 \frac{Z}{F}$ and $\mathbf{b}(\mathbf{u}) = (\frac{-12}{F} - \frac{6Z}{F^2}, 0, \frac{6}{F})^\top$, then $f(\mathbf{u}_r, \mathbf{u}_s) = \rho(\mathbf{x}_s - \mathbf{x}_r) + \tau(y_s - y_r)$, which corresponds to NTM.

Proof of Corollary 4: Inserting the specifications of $\mathbf{u}$, a and $\mathbf{b}$ in (M) yields

$$\begin{aligned} f(\mathbf{u}_r, \mathbf{u}_s) &= (c_5 + 2c_6^\lambda - 2c_6^\eta) (\ln F_s - \ln F_r) + 4c_8 (Y_s^2 - Y_r^2) + 2c_6^\eta \left(\frac{Z_s}{F_s} - \frac{Z_r}{F_r} \right) \\ &\quad + \left(\frac{c_7}{F_r} - \frac{2c_9Z_r}{F_r^2} \right) (F_s - F_r) + \left(-2c_6^\lambda - 8c_8Y_r \right) (Y_s - Y_r) + \frac{2c_9}{F_r} (Z_s - Z_r) \\ &= c_5 (y_s - y_r) + c_6^\lambda (v_s^\lambda - v_r^\lambda) + c_6^\eta (v_s^\eta - v_r^\eta) + c_7 (e^{y_s - y_r} - 1) \\ &\quad + c_8 (2Y_s - 2Y_r)^2 + c_9 \left(\frac{2Z_s}{F_s} - \frac{2Z_r}{F_r} \right) e^{y_s - y_r} = g(\mathbf{x}_s - \mathbf{x}_r). \quad \square \end{aligned}$$