Merchant Commodity Storage and Term Structure Model Error

Abstract
Merchants operations involves valuing and hedging the cash flows of commodity and energy conversion assets as real options based on stochastic models that inevitably embed model error. In this paper we quantify how empirically calibrated model errors concerning the futures term structure affect the valuation and hedging of natural gas storage. We find that even small futures price model errors – on the order of 1-2% of the empirical variance – can have a disproportionate impact on storage valuation and hedging. In particular, theoretically equivalent hedging strategies have very different sensitivities to model error, with one strategy exhibiting potentially catastrophic performance in the presence of these small errors. We propose approaches to mitigate the negative effect of term structure model error on hedging, also taking into account futures contract illiquidity, and provide theoretical justification for some of these approaches. Beyond commodity storage, our analysis should have relevance for other real and financial options that depend on futures term structure dynamics, as well as inventory, production, and capacity investment policies that rely on demand forecast term structures.

1. Introduction
Merchants manage commodity and energy conversion assets – copper mines, natural gas storage facilities and pipelines, power plants, and oil wells and refineries – as real options on commodity and energy prices. A canonical merchant operations problem involves buying or leasing commodity conversion assets and then using an operating policy to maximize their real option values by trading the underlying physical input and output commodities (Secomandi and Seppi 2014). Merchants also engage in financial hedging of these operational cash flows, to mitigate the effects to mismatches between financing payments and operational cash flows in the presence of credit frictions (see, e.g., Tirole 2006, §5.4 and references therein). This financial trading activity is analogous to how financial institutions hedge similar risks from their financial derivative trading operations (Hull 2010). These are complex multifunctional tasks faced every day in merchant operations.

The key components of merchant operations are a valuation model, an operating policy, and risk sensitivities for dynamic hedging. Each of these activities rely on an assumed model of the risk neutral dynamics of the underlying state variables (Smith and McCardle 1999, Birge 2000, Seppi 2002). Since, in practice, the risk neutral dynamics are not known with certainty, model errors can therefore cause merchants to misvalue commodity conversion assets and disrupt risk management (hedging) strategies.

Real options have been studied extensively in the theoretical and practice-based literature (Smith and McCardle 1999, Clewlow and Strickland 2000, Eydeland and Wolyniec 2003, Geman 2005, Leppard 2005), but model error has not been analyzed in the context of merchant operations. Our goal in this paper is thus to (i) quantify the impact of model error on real option valuation
and hedging in merchant operations and (ii) propose practical strategies to mitigate the exposure of merchant operations to model error.

We focus on commodity storage assets (lease contracts) that allow merchants to trade a commodity over time; specifically, we consider natural gas storage (Maragos 2002). Natural gas is an important commodity, representing 25% of total U.S. energy consumption in 2009 (EIA 2010), and is projected to become even more important in the future (EIA 2011a). Natural gas storage capacity amounts to roughly 17% of annual U.S. natural gas demand (EIA 2011b,c). At a price for natural gas of $4.9/mmBtu (in February 2014), the value of natural gas peak inventory is about $20B. Moreover, natural gas storage valuation has attracted a substantial amount of research (Chen and Forsyth 2007, Boogert and de Jong 2008, Thompson et al. 2009, Carmona and Ludkovski 2010, Lai et al. 2010, Secomandi 2010b, Bjerksund et al. 2011, Boogert and de Jong 2011/12, Thompson 2012, and Wu et al. 2012).

As described in the practice-based literature (Maragos 2002, pp. 440 and 449-453, and Gray and Khandelwal 2004a,b), natural gas storage merchants use multifactor models for the term structure of commodity futures prices (Clewlow and Strickland 2000, Chapter 8, Eydeland and Wolyniec 2003, Chapter 5 and pp. 351-367, and Geman 2005, Chapter 3). Empirically, the dynamics of commodity and energy futures price term structures are dominated by a few large common factors. A typical number of factors is three, which have qualitative interpretations of level (shifting), slope (tilting), and curvature (bending) effects (Cortazar and Schwartz 1994, Clewlow and Strickland 2000, Chapter 8, Blanco et al. 2002, Tolmasky and Hindanov 2002, Geman and Nguyen 2005, Borovkova and Geman 2008, and Frestad 2008). However, fewer factors (Manoliu and Tompaidis 2002, Borovkova and Geman 2006, Suenaga et al. 2008, and Wu et al. 2012) or more factors (Eydeland and Wolyniec 2003, pp. 351-367, Gray and Khandelwal 2004a,b, Bjerksund et al. 2011, and Thompson 2012, 2013) are also used. In particular, Eydeland and Wolyniec (2003, pp. 351-367) and Gray and Khandelwal (2004a,b) discuss modeling the full term structure of the natural gas futures curve using as many common factors as there are futures delivery dates over the term of a storage contract, an approach related to string and BGM models used to value fixed income options (Kennedy 1994, Brace et al. 1997, Longstaff et al. 2001). Moreover, Frestad (2008) and Suenaga et al. (2008) model maturity specific (idiosyncratic) factors in addition to common factors.

There is an infinite variety of possible term structure model errors, so we look to empirical data to impose realistic structure on the types of model errors we analyze. In other words, we study the impact of empirically realistic model errors that merchant storage traders are likely to face. We calibrate a flexible family of multifactor futures curve models (Cortazar and Schwartz 1994, Blanco
et al. 2002). For New York Mercantile Exchange (NYMEX) natural gas futures prices, as expected, we find that the three most important common factors explain about 98-99% of the observed daily variance of log futures price returns. However, given such empirical evidence, merchant traders are still uncertain about the correct model for natural gas futures term structure dynamics. For example, does the treatment of the last 1-2% of observed log futures price return variance matter for storage valuation and hedging? We thus center our model error analysis on various term structure models that differ in their treatment of the three most important common factors and this small residual variance.

Our quantitative analysis is based on applying the rolling intrinsic (RI) operating policy, which is both widely used in practice and near optimal (see Lai et al. 2010 and references therein), to value storage contracts with twenty-four month terms. We analyze versions of practical hedging methods based on both cash flow replication (delta hedging; Hull 2010, Chapter 6, and Luenberger 2014, §12.5) and minimization of hedged cash flow variance (Luenberger 2014, §12.10). In particular, for delta hedging we consider bucket hedging (BH; Driessen et al. 2003), which trades futures contracts with delivery dates corresponding to each date on which operational trading has a cash flow, and factor hedging (FH; Cortazar and Schwartz 1994 and Clewlow and Strickland 2000, §9.5), which just trades as many futures contracts as there are factors in the assumed futures price model. FH is appealing because, in contrast to BH, it is feasible when there are fewer futures contracts available than cash flow dates for the option. Both of these methods take the storage deltas as inputs. We estimate these deltas by extending the pathwise delta estimation approach (see, e.g., Glasserman 2004, §7.2 and references therein), under some technical conditions that we show hold for an optimal inventory policy under a mild parameter assumption (Secomandi and Wang 2012 apply the pathwise approach to real options without inventory). We also obtain a novel expression for the FH positions.

Our analysis shows that model errors concerning even small amounts of empirical variance can have a disproportionate impact on storage valuation and hedging. The inclusion/exclusion/treatment of the last 1-2% of the empirical log futures price return variance can change storage valuations by ±14%, an insight that is relevant for merchants when deciding how to bid for acquiring the capacity of storage assets. The reason for this numerical disparity is that storage valuation depends strongly on the ability to trade on nonpersistent price shocks, even if they are statistically small in comparison to statistically large but more persistent shocks. The impact of model error on hedging performance, which we measure as the reduction in the variance of the operational cash flows using both simulated and historical data, can be even more dramatic. The performance of hedges
that in the absence of model error are theoretically equivalent can be orders of magnitude different
given small model errors pertaining, again, to just 1-2% of the log futures price return variance.
Specifically, we find that BH performs near-optimally and is remarkably robust to model error, an
observation for which we provide some theoretical support. In contrast, FH is very sensitive to
which futures contracts are used to hedge and the sizes of the positions taken. Implementing this
approach using the nearby, and most liquid, futures contracts, a method that we label naïve FH
(NFH), performs disastrously in the presence of even small model errors. The differential perfor-
mane of these hedging strategies is closely tied to the size of the futures positions taken. Large
position sizes can magnify hedging errors due to even small amounts of futures curve randomness
mistakenly omitted from the price model.

We also focus on mitigating the effect of term structure model error on hedging. We develop
a novel FH method, fine-tuned FH (FTFH), which, by (heuristically) minimizing the size of the
trading positions, performs near optimally. However, when the storage contract is long-dated,
both BH and FTFH are infeasible due to futures contract illiquidity. We thus propose a novel
minimum variance method, constrained minimum variance hedging (CMVH), which trades only
liquid contracts (see also Frestad 2012). We find that CMVH performs quite well in the presence
of both model error and futures contract illiquidity.

Our research improves our understanding of the effect of realistic term structure model errors
on merchant commodity storage management, and provides methods to mitigate the negative effect
on hedging of such errors. More broadly, our work has potential relevance for valuing and hedging
other commodity conversion assets and also real and financial options using multifactor models,
which are likely specified with error in practice. Examples include commodity processing, refining,
and transport assets (Secomandi 2010a, Wu and Chen 2010, Boyabatli et al. 2011, Lai et al. 2011,
commodity swing and Bermudan options (Jaillet et al. 2004 and Detemple 2006), and mortgages
and interest rate caps, floors, and swaptions (Longstaff et al. 2001, Fan et al. 2001, Driessen et al.
2003, and Veronesi 2010). Model error is also likely to arise in practice when applying the martingale
model of demand forecast evolution (Graves et al. 1986 and Heath and Jackson 1994) to determine
inventory/production management and capacity investment policies under demand uncertainty, for
which financial hedging can be relevant (Birge 2000, Van Mieghem 2003, Berling and Rosling 2005,
Gaur and Seshadri 2005, Caldentey and Haugh 2006, Ding et al. 2007, Goel and Gutierrez 2011,

We present a model of futures price dynamics in §2. We discuss the valuation and hedging of
natural gas storage in the absence of model error in §3 and §4, respectively. We calibrate our futures price model in §5. In §6 we use the results of this calibration to introduce term structure model error and discuss its potential impact on storage valuation and hedging. We propose approaches to mitigate the negative effect of model error on hedging in §7. We report on our numerical analysis in §8. Section 9 concludes. Online Appendix A extends the pathwise delta estimation analysis of §4.2 to the case of optimal storage policies. Online Appendix B includes proofs. Online Appendix C presents additional results in support of our numerical hedging analysis.

2. A Family of Futures Price Evolution Models

Multifactor models are often used to describe the risk-neutral dynamics of the term structure of commodity futures prices. Consider the first $N > 1$ contracts in the natural gas futures curve over the term $[T_0, T_{N-1}]$ of a storage contract, with $T_0$ set equal to 0. We denote by $F(t, T_m)$ the futures price at time $t \in [0, T_m]$ with maturity date $T_m$, for all $m$ in the set $N$ of maturity labels $\{0, \ldots, N-1\}$. The spot price at time $T_m$ is $F(T_m, T_m)$. For a fixed number of common factors $K \in \{1, \ldots, N-1\}$, the following stochastic differential equation describes the risk-neutral dynamics of $F(t, T_m)$ during the time interval $[0, T_m]$:

$$
\frac{dF(t, T_m)}{F(t, T_m)} = \sum_{k \in K} \sigma_{m,k}(t)dZ_k(t),
$$

where $K$ is the set $\{0, \ldots, K-1\}$ of factor labels, $\sigma_{m,k}(t)$ is the loading coefficient for factor $k$ of the futures price with maturity date $T_m$ at time $t$, $dZ_k(t)$ is a standard Brownian motion increment corresponding to factor $k$, and the standard Brownian motion increments for the different factors are all uncorrelated. Seasonality in spot price levels is captured by the initial futures prices themselves. Model (1) also allows for possible seasonality in the futures log return covariance matrix, because the factor loading coefficients can depend on calendar time $t$.

Factor models of futures price dynamics are mathematically equivalent to factor models of spot price dynamics (Cortazar and Schwartz 1994 and Secomandi and Seppi 2014, Chapters 3 and 4). For example, the two-factor spot price model of Schwartz and Smith (2000) and Pilipovic (2007, §5.3.2.2, §6.7) is equivalent to a two-factor futures curve model with parallel shift loadings on one factor and decaying loadings on a second (correlated) factor (see Ross 1997, Schwartz 1997, Jaillet et al. 2004, and Casassus and Collin-Dufresne 2005 for related models). We conduct our analysis using a dynamic futures curve specification both because of its empirical flexibility and because, in practice, futures are widely used for hedging (Clewlow and Strickland 2000, §8.5 and Chapter 9).

Our specification of model (1) restricts each coefficient $\sigma_{m,k}(\cdot)$ to be a constant $\sigma_{m,k,n}$ over each time interval $[T_n, T_{n+1})$, with $n \in N$ and $m$ in the set $N_{n+1}$ of residual maturity labels.
\{n+1, \ldots, N-1\}. That is, distinct factor loadings \(\sigma_{m,k,n}\) are obtained for each futures maturity for each trading month. This seasonal specification captures the important role of weather in natural gas supply and demand. Given \(n\) and \(m \in N_{n+1}\), under this specification the dynamics of the futures price \(F(t, T_m)\) during the time interval \([T_n, T_{n+1}]\) are

\[
\frac{dF(t, T_m)}{F(t, T_m)} = \sum_{k \in K} \sigma_{m,k,n} dZ_k(t).
\] (2)

3. Valuation

This section is based in part on Lai et al. (2010, §2, §§3.2-3.3). Natural gas storage contracts give merchants the right to inject, store, and withdraw natural gas at a storage facility over a finite time horizon, subject to capacity and inventory constraints. Capacity flow constraints specify the maximal amount of natural gas, measured in million British thermal units (mmBtus), which a merchant can inject or withdraw per unit of time. Inventory constraints specify the minimal and maximal amounts of natural gas inventory that the merchant can hold at any given point in time. A contract also specifies injection and withdrawal charges, and fuel losses.

We assume that the contracted storage facility is located nearby a liquid wholesale spot market, in which the merchant executes its physical trading. This is realistic for North America (which has roughly one hundred geographically dispersed wholesale markets for natural gas) and for the United Kingdom (the National Balancing Point), Belgium (Zeebrugge), Germany (Emden), and the Netherlands (the Title Transfer Facility). If the storage facility is not located at a market hub, then transport capacity and cost for the pipeline connecting the storage facility to the relevant hub must also be included. We also assume that a natural gas futures market is associated with the wholesale physical market. This is again realistic for North America where NYMEX and ICE trade futures contracts with delivery at Henry Hub, Louisiana and basis swaps for about forty other locations (financially settled forward locational price differences relative to Henry Hub). In the United Kingdom, ICE trades natural futures associated with the National Balancing Point. The European Energy Exchange also trades natural gas futures.

We are interested in valuing the cash flows associated with making physical natural gas trading decisions on a monthly basis. This is realistic because the spot market in the United States is most liquid during the monthly “bid week” in which blocks of gas for the ensuing month are traded. The timing of the physical trading/inventory decisions is assumed to coincide with the futures maturity dates, which define the stage set. Let \(a\) denote an action. A purchase-and-injection is a negative action (i.e., it generates a negative cash flow), a withdrawal-and-sale is a positive action, and the
do-nothing action is zero. Natural gas purchased (respectively, sold) at time $T_n$ is available in (respectively, removed from) storage by time $T_{n+1}$. A nonzero action thus represents a commodity flow in between two successive stages.

The minimal and maximal inventory levels are 0 and $\bar{x} \in \mathbb{R}_+$, respectively, making the set of feasible inventory levels $[0, \bar{x}] =: \mathcal{X}$. The injection and withdrawal capacities per stage are $C^I < 0$ and $C^W > 0$, respectively. The feasible injection action, withdrawal action, and action sets, respectively, with feasible inventory $x$ are $\mathcal{A}^I(x) := [C^I \vee (x - \bar{x}), 0]$, $\mathcal{A}^W(x) := [0, x \wedge C^W]$, and $\mathcal{A}(x) := \mathcal{A}^I(x) \cup \mathcal{A}^W(x) \{ \cdot \wedge \cdot \equiv \min\{\cdot, \cdot\} \text{ and } \cdot \vee \cdot \equiv \max\{\cdot, \cdot\}\}$. We model in-kind fuel losses using the coefficients $\phi^W \in (0, 1]$ and $\phi^I \geq 1$ for withdrawals and injections, respectively. The marginal withdrawal and injection costs are $c^W$ and $c^I$, respectively. Thus, given an action $a$ and the spot price $s$, the per stage cash flow function $p(a, s)$ is $(\phi^I s + c^I) a$ if $a \in \mathbb{R}_-$, 0 if $a = 0$, and $(\phi^W s - c^W) a$ if $a \in \mathbb{R}_+$.

The storage contract is managed using a feasible physical trading policy $\pi$. Using the more compact notation $F_{n,m}$ for the futures price $F(T_n, T_m)$, let $F_n := (F_{n,m}, m \in \mathcal{N}_n)$, $\forall n \in \mathcal{N}$, and $F'_n := (F_{n,m}, m \in \mathcal{N}_{n+1})$, $\forall n \in \mathcal{N} \setminus \{N - 1\}$, denote the futures curves at time $T_n$ inclusive and exclusive, respectively, of the spot price, $s_n \equiv F_{n,n}$, with the conventions $F_N := 0$ and $F'_{N-1} := 0$. Let $A_n^\pi(x, F_n)$ be the decision rule giving the amount of gas sold or bought under a policy $\pi$ at stage $n$ in state $(x, F_n)$ and $x_n^\pi$ the inventory level reached at stage $n$ by policy $\pi$. Denote risk-neutral expectation by $\mathbb{E}$ and the per stage risk-free discount factor by $\delta$. The value function corresponding to policy $\pi$ in state $(x, F_n)$ at stage $n$ is

$$V_n^\pi(x, F_n) := \sum_{m=n}^{N-1} \delta^{m-n-1} \mathbb{E}[p(A_m^\pi(x_m^\pi, F_m), s_m)|x, F_n].$$ (3)

The optimal storage valuation problem can be formulated as a Markov decision process, but this model is intractable due to the curse of dimensionality as decisions are conditioned on the entire futures curve $F_n$ at each stage (see Appendix A for a stochastic dynamic programming formulation for this model). As a result, in practice natural gas storage optionality is often managed using simpler heuristic operating policies. Our theoretical analysis of hedging in §4 considers feasible operating policies; our numerical analysis in §8 assumes use of the RI heuristic operating policy, which is both widely used among practitioners and has been observed by Lai et al. (2010) to be nearly optimal (see also Thompson 2012 and Wu et al. 2012). This policy is based on sequential reoptimization of the intrinsic value model (i.e., the deterministic equivalent of the stochastic dynamic program for optimal storage valuation in Online Appendix A). The value function of the RI policy can be estimated by a Monte Carlo simulation of the evolution of the futures curve. We
refer the reader to Lai et al. (2010, §§3.2-3.3) for details on this policy. We expect that our model error analysis in §8 would carry over qualitatively to other operating policies.

4. Hedging

Under the price model specifications presented in §2 hedging occurs, in theory, in continuous time; therefore, we map the discrete-time function $V_n^\pi(\cdot, \cdot)$ in (3) into a continuous-time value function. On a date $t \in [T_{n-1}, T_n]$ for $n \in N \setminus \{0\}$, we denote as $G_n(t)$ the vector of futures prices $(F(t, T_m), m \in N_n)$ (at time $t = T_n$ this vector includes the spot price $s_n$). For every $n \in N \setminus \{0\}$, $t \in [T_{n-1}, T_n]$, and state $(x, G_n(t)) \in X \times \mathbb{R}^{N-n}$, we define the continuous-time value function for a given feasible operating policy $\pi$ as

$$U_n^\pi(t, x, G_n(t)) := \tilde{\delta}(t, T_n) \mathbb{E}[V_n^\pi(x, F_n) | G_n(t)],$$  

(4)

where $\tilde{\delta}(t, T_n)$ is the risk-free discount factor from time $T_n$ back to time $t$ ($\tilde{\delta}(T_{n-1}, T_n)$ corresponds to the constant one-period risk-free discount factor $\delta$). The function $U_n^\pi(t, x, G_n(t))$ is the time $t$ value of having $x$ inventory units in storage at time $T_n$ given the vector $G_n(t)$ of futures prices at time $t$ when using policy $\pi$ to manage storage.

For hedging, we need to know how the value of storage changes in response to changes in the futures curve. We focus on price model (2). By Itô’s lemma (see, e.g., (Glasserman 2004, Theorem B.1.1)), the dynamics of $U_n^\pi(t, x, G_n(t))$ at time $t$ in $(T_{n-1}, T_n]$ are

$$dU_n^\pi(t, x, G_n(t)) = D(t)dt + \sum_{m \in N_n} \Delta_{n,m}^\pi(x, t, G_n(t))dF(t, T_m),$$

(5)

where the drift is

$$D(t) := \frac{\partial U_n(t, x, G_n(t))}{\partial t} + \frac{1}{2} \sum_{m \in N_n} \sum_{j \in N_n} \frac{\partial^2 U_n(t, x, G_n(t))}{\partial F(t, T_m) \partial F(t, T_j)} \sum_{k \in K} \sigma_{m,k,n-1}F(t, T_m)\sigma_{j,k,n-1}F(t, T_j)$$

and the storage delta with respect to the time $t$ futures prices with maturity $T_m$, with $m \in N_n$, is

$$\Delta_{n,m}^\pi(t, x, G_n(t)) := \frac{\partial U_n^\pi(t, x, G_n(t))}{\partial F(t, T_m)}.$$  

(6)

When $t = T_{n-1}$, the inventory $x$ is held fixed, because the deltas are computed right after the commercial implementation of a feasible action at time $T_{n-1}$. Otherwise, when $t \in (T_{n-1}, T_n)$, inventory is fixed based on the prior decision at $T_{n-1}$.

4.1 Replication

To ease exposition, we mostly simplify the suffix $(t, x, G_n(t))$ to $(t)$. Denote by $q_{n,m}^\pi(t)$ the replicating position corresponding to the futures contract with maturity $T_m$, $m \in N_n$, at time $t$ when hedging $U_n^\pi(t)$. The dynamics of the value $\Pi_n^\pi(t)$ of a portfolio that is long physical storage and
continuously *shorts* these futures positions are $d\Pi_n^\pi(t) = dU_n^\pi(t) - \sum_{m \in \mathcal{N}_n} q_n^{\pi,m}(t) dF(t, T_m)$, which, using (5), under our specification (2), can be expressed as

$$d\Pi_n^\pi(t) = D(t) dt + \sum_{k \in K} \left[ \sum_{m \in \mathcal{N}_n} \Delta_n^{\pi,m}(t) F(t, T_m) \sigma_{m,k,n-1} - \sum_{m \in \mathcal{N}_n} q_n^{\pi,m}(t) F(t, T_m) \sigma_{m,k,n-1} \right] dZ_k(t).$$

(7)

It follows from (7) that the stochastic variability is hedged when

$$\sum_{m \in \mathcal{N}_n} q_n^{\pi,m}(t) F(t, T_m) \sigma_{m,k,n-1} = \sum_{m \in \mathcal{N}_n} \Delta_n^{\pi,m}(t) F(t, T_m) \sigma_{m,k,n-1}, \forall k \in K.$$  

(8)

This is a system of $K$ linear equations (one for each factor) with $N - n$ unknowns (the futures positions $q_n^{\pi,m}(t)$). When the number of factors driving the futures curve is equal to or greater than the number of futures contracts (i.e., $K \geq N - n$), setting $q_n^{\pi,m}(t) = \Delta_n^{\pi,m}(t)$ for all $m \in \mathcal{N}_n$ solves the system of linear equations (8). Otherwise (i.e., $K < N - n$), this system of linear equations is underdetermined and multiple replicating positions exist. We consider several different particular solutions.

**BH.** BH sets the replicating positions equal to the deltas, i.e., $q_n^{\pi,m}(t) = \Delta_n^{\pi,m}(t)$ for all $m \in \mathcal{N}_n$, which provides a solution to the system of linear equations (8). BH does not depend directly on the factor loadings $\sigma_{m,k,n-1}$; rather the futures curve factor structure only affects bucket hedging indirectly through the deltas. As a result, the bucket hedges for different futures price models with different numbers of factors differ only to the extent that the corresponding deltas differ.

**FH and NFH.** FH only takes positions in as many futures contracts as the number of futures curve factors, $K$. Consequently, FH relies strongly on the assumed futures factor structure. In principle, any $K$ contracts can be used as long as their factor loadings are linearly independent. We include in set $\mathcal{H}_n$ the maturity labels of the $K$ futures contracts used for hedging, and collect the remaining unused maturity labels in set $\mathcal{H}_n^c = \mathcal{N}_n \setminus \mathcal{H}_n$. We set $q_n^{\pi,m}(t) := 0$ for all $m \in \mathcal{H}_n^c$ and then include in the column vector $q_n^{\pi,\mathcal{H}_n}(t)$ the positions of the contracts with labels in set $\mathcal{H}_n$ obtained as the solution to the remaining system of $K$ linear equations in $K$ unknowns from (8). The solution, given in Proposition 1, uses the following notation: At time $t \in [T_{n-1}, T_n)$, $\text{diag}(G^\mathcal{H}_n(t))$ and $\text{diag}(G^{\mathcal{H}_n^c}(t))$ are diagonal matrices corresponding to the time $t$ futures prices in sets $\mathcal{H}_n$ and $\mathcal{H}_n^c$; $\Delta_n^{\pi,\mathcal{H}_n}(t)$ and $\Delta_n^{\pi,\mathcal{H}_n^c}(t)$ are column vectors of deltas, which together form $\Delta_n^\pi(t)$, corresponding to maturities labels in sets $\mathcal{H}_n$ and $\mathcal{H}_n^c$; and $B_n$ and $E_n$ are $K \times K$ and $(N - n - K) \times K$ submatrices of factor loading coefficients for maturity labels in sets $\mathcal{H}_n$ and $\mathcal{H}_n^c$. We denote transposition by the superscript $^T$. 

9
Proposition 1 (FH positions). Suppose that $B_{n-1}$ is invertible. Pick $n \in \mathcal{N}$ such that $N-n > K$ and $t \in [T_{n-1}, T_n)$. The replicating FH positions in futures contracts corresponding to set $\mathcal{H}_n$ are

$$q^T \mathcal{H}_n(t) = \Delta_n^\pi \mathcal{H}_n(t) + \text{diag}^{-1}(G^\mathcal{H}_n(t)) \left( B_{n-1}^T \right)^{-1} E_{n-1}^T \text{diag}(G^\mathcal{H}_n(t)) \Delta_n^\pi \mathcal{H}_n(t).$$

(9)

Each position in a traded futures is the sum of the delta for that futures contract and a linear combination of the deltas corresponding to the untraded futures contracts. In contrast to BH, FH depends on the futures curve factor structure directly via the factor loadings submatrices $B_{n-1}$ and $E_{n-1}$ as well as indirectly via the deltas.

Different implementations of FH use different subsets of futures to delta hedge. We consider NFH as our benchmark FH approach. At time $t \in [T_{n-1}, T_n)$ NFH simply takes positions in the $K$ shortest maturity futures contracts (i.e., it sets $\mathcal{H}_n$ equal to $\{n, \ldots, n + K - 1\}$). As documented in §5, these are typically the most liquid contracts. Thus, NFH is a natural FH implementation. However, as discussed in §7.1, the performance of NFH may be fragile in the presence of model error. We present FTFH in §7.1 to rectify this problem.

4.2 Estimating Deltas

The storage deltas are critical inputs to both BH and FH. Typically, they must be estimated numerically. Assumption 1 gives a set of sufficient conditions that allows us to extend to storage the pathwise approach (see, e.g., Glasserman 2004, §7.2) for Monte Carlo unbiased deltas estimation.

Assumption 1 (Lipschitz continuity and derivative characterization). (a) In every stage $n \in \mathcal{N}$, for each given inventory $x \in \mathcal{X}$ the function $V_n^\pi(x, F_n)$ is Lipschitz continuous in the futures curve $F_n \in \mathbb{R}^{N-n}$; that is, there exists $L_n^\pi(x) \in \mathbb{R}_+$ such that $|V_n^\pi(x, F_n^2) - V_n^\pi(x, F_n^1)| \leq L_n^\pi(x) \sum_{m=n}^{N} |F_{n,m}^2 - F_{n,m}^1|$, for all $F_n^1, F_n^2 \in \mathbb{R}^{N-n}$. (b) Moreover, at every futures curve $\overline{F}_n$ where $V_n^\pi(x, F_n)$ is differentiable with respect to each futures price in $F_n$, the decision rule $A_n^\pi(x, F_n)$ given the policy $\pi$ has a unique action denoted by $a_n^\pi(x, F_n)$ and for all $m \in \mathcal{N}_n$

$$\frac{\partial V_n^\pi(x_n, F_n)}{\partial F_{n,m}} \bigg|_{F_n = \overline{F}_n} = \frac{\partial p(a_n^\pi(x_n, F_n), s_n)}{\partial F_{n,m}} \bigg|_{s_n = \overline{F}_n} + \frac{\partial \mathbb{E}[V_{n+1}^\pi(x_n - a_n^\pi(x_n, F_n), F_{n+1})]}{\partial F_{n,m}} \bigg|_{F_n = \overline{F}_n}.$$  

The second part of this assumption roughly states that the feasible policy $\pi$ has constant actions in neighborhoods of differentiability of its value function with respect to each element of the futures curve. As shown in Online Appendix A, under a mild assumption, an optimal operating policy satisfies Assumption 1. We denote by $1\{\mathcal{E}\}$ the indicator function of event $\mathcal{E}$.

Proposition 2 (Pathwise deltas). Under Assumption 1, for every $n \in \mathcal{N} \setminus \{0\}$ it holds that

$$\Delta_n^\pi(t, x_n, G_n(t)) = \frac{\overline{G}(t, T_m)}{G(t, T_m)} \mathbb{E}[\phi W_1 \{A_n^\pi(x_m, F_m) < 0\} + \phi W_1 \{A_n^\pi(x_m, F_m) > 0\} | x_n, G_n(t)]$$

(10)
for all \( m \in \mathcal{N}_n, \ x_n \in \mathcal{X}, \ t \in [T_{n-1}, T_n), \) and \( G_n(t) \in \mathbb{R}^{N-n}. \)

Expression (10) can be used to estimate deltas by Monte Carlo simulation simultaneously with the valuation of a policy \( \pi. \) In \( \S 8 \) we use this expression with the RI policy, even though the corresponding value function can violate the Lipschitz condition in Assumption 1. This calculation may yield biased delta estimates but can be implemented efficiently. Our computational results in \( \S 8 \) suggest that any bias in our calculated deltas is small. This is expected given that Lai et al. (2010) find that the RI policy is near optimal (see also Thompson 2012, Wu et al. 2012). This approach seems better than calculating deltas by resimulation, which is both biased and computationally expensive (Glasserman 2004, \S 7.1).

The term on the right hand side of (10) can be interpreted as being proportional to a “weighted average” of the action taken by a feasible policy in a given stage and state. To see this, notice that the ratio \( s_m/F(t, T_m) \) is a nonnegative random variable with mean one (because \( \mathbb{E}[s_m|F(t, T_m)] = F(t, T_m) \)). This ratio weighs the amount of commodity traded by policy \( \pi \) in state \((x_n, F_m)\) at stage \( n, \) i.e., \( (\phi I \{ A_m(x_n, F_m) < 0 \} + \phi W \{ A_m(x_n, F_m) > 0 \}) A_m(x_n, F_m) \). Consequently, the expression \( \mathbb{E} [(\phi I \{ A_m(x_n, F_m) < 0 \} + \phi W \{ A_m(x_n, F_m) > 0 \})s_m A_m(x_n, F_m)/F(t, T_m) | x_n, G_n(t)] \) is a weighted expectation of the amount of commodity traded by policy \( \pi \) in this state and stage. The delta \( \Delta_{n,m}^\pi (t, x, G_n(t)) \) is proportional to this expectation according to \( \delta(t, T_m). \)

**Proposition 3 (Bounds on deltas).** Under Assumption 1, for every \( n \in \mathcal{N} \setminus \{0\} \) it holds that \( \delta(t, T_m) \phi C^I \leq \Delta_{n,m}^\pi (t, x, G_n(t)) \leq \delta(t, T_m) \phi W C^W, \) for all \( m \in \mathcal{N}_n, \ x_n \in \mathcal{X}, \ t \in [T_{n-1}, T_n), \) and \( G_n(t) \in \mathbb{R}^{N-n}. \)

Proposition 3 lets us relate the size of the RI policy deltas that we compute in \( \S \S 8.2.2-8.2.3 \) to those of a policy that does satisfy Assumption 1, in particular an optimal policy (under a mild assumption satisfied in our numerical analysis).

5. **Calibration**

Consider a merchant trader who wants to implement storage valuation and hedging based on a futures price term structure model. Consistent with industry practice (Clewlow and Strickland 2000, \S 8.6, Blanco et al. 2002), using twelve separate monthly Principle Component Analyses (PCAs) is a natural approach to estimate the constant intra-month factor loadings of the futures price evolution model (2). We implement this estimation using daily NYMEX natural gas futures prices from January 1997 to December 2012.

Figure 1 displays the monthly loading coefficients for twenty-three monthly maturities for the first four factors. We distinguish between months in the heating season (November-March), panels
Figure 1: The estimated loading coefficients of the first four factors for maturities one to twenty-three from month-specific PCAs over January 1997 to December 2012.

(a), (c), (e), and (g), and the rest of the year, panels (b), (d), (f), and (h). Typically, the factor loading coefficients decrease rapidly, with the coefficients of a given factor being less than half of those of the previous factor. The first factor changes the slope of the natural gas futures term structure. In particular, the first factor induces positively correlated shocks along the futures term structure (since the coefficients are of the same sign), but it moves the short end of the term structure more than the long end (since the coefficients decline in magnitude in time to delivery). The second factor also shocks the slope of the futures term structure, but the changes at the long and short ends of the curve are negatively correlated (since the coefficients change sign). The third factor shocks the curvature of the futures term structure (since the signs of short-term and long-term futures coefficients have the opposite sign from the intermediate term coefficients). The
fourth factor is typical of the remaining PCA factors. They induce small irregular “squiggles” along the futures term structure (since the signs of the coefficients change multiple times). In terms of seasonality, the shapes of the loading coefficient curves do appear to change across different months.

Figure 2 displays the percentages of the total variance in the observed log price changes explained by the first four factors for each of the twelve monthly PCAs. The first factor captures roughly 88.7-93.5% of this variance, the second factor 5.0-7.6%, the third factor 0.6-3.0%, and the fourth factor always less than 1%. The first three factors explain just over 98% of the total variance in the observed log price changes in each month.

Our calibration ignores liquidity concerns because we only have available closing futures prices. Figure 3 displays the average trading volume of the natural gas futures contracts for the first twenty-three maturities from January 1997 to December 2012. As is apparent from this figure, futures with distant maturities suffer from lower liquidity, which could result in high transaction costs (wide bid-ask spreads). Although we do not consider this issue in our calibration, we model the impact of limited liquidity on hedging in §6.

6. Term Structure Model Error and Its Potential Impact

The PCA estimation results in §5 are the starting point for our analysis of empirically likely model error. Merchant storage traders use PCA estimations to calibrate an assumed model of futures curve dynamics on which they then base their valuation, operations, and hedging decisions. However, they know that this approach contains a variety of likely model errors. First, there is parameter error
in the estimated factor loading coefficients. Second, it may be unclear how to treat the estimated PCA factors. If estimated randomness is excluded from the assumed futures price model, then the model may have missing factors (if bona fide factors are excluded) or missing noise (if the excluded randomness is really due to maturity-specific idiosyncratic shocks rather than common factors). If instead estimated factors that are not bona fide are included, the assumed model includes spurious factor error. Third, the estimated PCA may be structurally inconsistent with the true process due to omitted jumps, stochastic volatility, or incorrect functional specifications for the factor loadings. These model errors have potential negative effects on both storage valuation and hedging.

Model error in the assumed futures term structure model can affect the storage valuation through two channels. First, the operating policy may depend on incorrect beliefs about the correct futures curve dynamics, even if optimal for the assumed futures curve dynamics. Second, the expectation of the storage cash flows in the valuation may be under an incorrect assumed risk neutral measure. The incorrect operating policy effect of valuation model error is not relevant when gas storage is operated using the RI operating policy, which depends on the observed futures curve, but the incorrect expectations effect is.

Term structure model error can also affect storage hedging performance in two ways. First, the deltas based on the assumed futures curve model may be different from the deltas corresponding to the correct futures model. Second, futures positions obtained for the assumed futures curve dynamics may not perform well given the correct dynamics. Different hedging strategies might have different sensitivities to these model errors. For example, as discussed in §4.1, BH does not depend directly on the factor structure of the futures curve dynamics. As a result, term structure model error only affects the BH performance through the storage deltas. In contrast, FH relies on
the futures curve factor structure to determine the replicating positions, and, thus, is likely to be more susceptible to term structure model error.

To analyze in a controlled setting the potential effect of term structure model error on storage valuation and hedging when a merchant uses a futures price model such as (2), we consider two empirically plausible families of futures price models based on our PCA estimation in §5. The first family consists of common factor models (2) in which different numbers of the PCA common factors are considered to be bona fide and the rest of the estimated factors are viewed as spurious, and are hence ignored. The second family consists of common factor plus noise models, specified below, with both different numbers of PCA common factors and maturity-specific (idiosyncratic) noise factors that capture the empirical variability associated with the excluded common factors. In the first case, the modeled log futures price return variances are less than the empirical variances (unless all factors are included, in which case the modeled and empirical variances coincide). In the second case, the modeled and empirical variances are equal, but the covariance structures differ.

Specifically, the common factor plus noise models add maturity-specific (idiosyncratic noise) factors with loading coefficients $\hat{\sigma}_{m,n}$ to price model (2), so that the dynamics of the futures price $F(t,T_m)$ during the time interval $[T_n, T_{n+1})$ become

$$
\frac{dF(t,T_m)}{F(t,T_m)} = \sum_{k \in \mathcal{K}} \sigma_{m,k,n} dZ_k(t) + \hat{\sigma}_{m,n} d\tilde{Z}_m(t),
$$

where $d\tilde{Z}_m(t)$ is a standard Brownian motion increment specific to maturity $T_m$ uncorrelated with every other maturity specific standard Brownian motion increment and each of the common standard Brownian motion increments $dZ_k(t)$. We calibrate the loading coefficients $\hat{\sigma}_{m,n}$ by matching the residual log futures price returns variance for each maturity unexplained by the chosen common factors.

Given the PCA factor sizes in Figure 2, a natural form of model uncertainty concerns the statistically small residual common factors after the first three (standard) large factors are included. Our investigation of model error thus posits that a merchant uses an assumed $K$ common factor model while the actual futures prices are generated by one of the following three hypothetically true reference models: (i) A 3 common factor model (that captures close to 99% of the empirical log futures price returns variance); (ii) an $(N - 1)$ common factor model (that takes the full empirical variance as true); and (iii) a 3 common factor plus noise model (that attributes to uncorrelated noise the about 1-2% PCA residual variance).

The template in Figure 4 displays the resulting term structure model errors, which correspond to different combinations of omitted or spurious factors and missing noise. Here, DGNF denotes
the hypothetically true futures price data generating number of (common) factors. For example, using an assumed common factor model with $K$ equal to 1 when the hypothetically true model has DGNF equal to 3 results in model error due to two missing large common factors, while assuming $K$ equal to 4 when the hypothetically true model has DGNF equal to 3 and includes noise results in model error due to the inclusion of one spurious common factor and missing noise.

Our model error template allows us to investigate through a series of what if experiments the likely potential impact of different model errors on natural gas storage valuation and hedging using Monte Carlo simulation. That is, we can estimate the value of storage by simulating sample paths of futures prices from the hypothetically true models in our template, and compare the resulting storage valuations obtained under various assumed common factor models. Similarly, we can apply various hedging strategies based on different assumed common factor models, and measure their performance on futures curve paths generated by different hypothetically true models. We perform this valuation and hedging model error analysis in §8.1 and §8.2.2, respectively. We also measure the effect of real model error on merchant natural gas storage hedging in §8.2.3 using real world futures curve paths.

The low liquidity for futures with long-dated maturities discussed in §5 can limit the availability of these contracts for hedging creating an important friction for hedging in practice. In our analysis we allow for the possibility that extremely illiquid futures cannot be traded, and refer to this friction as the missing contracts friction. This friction interacts with model error because some hedging strategies may become infeasible when they require trading in illiquid contracts. In §7.3 we propose CMVH to manage the missing contracts friction.

7. Coping with Term Structure Model Error in Hedging

In this section we propose practical approaches to mitigate the negative effect of term structure model error on hedging, also taking into account the missing contracts friction. We propose FTFH in §7.1, discuss the relationship between BH and MVH in §7.2, and present CMVH in §7.3.
7.1 FTFH
The size of the futures trading positions is, in theory, irrelevant in the absence of model error. Size becomes important, however, if large trading positions amplify unmodeled randomness in the realized price changes. Hence, we expect hedges with smaller positions to be more robust to these model errors. This issue of size is only relevant for FH because the BH positions are given.

NFH ignores the issue of size, i.e., it makes no attempt to control the size of the trading positions, and may take extreme positions. We propose FTFH to control the size of factor hedging trading positions. Since determining the set of contracts with the smallest possible FH positions requires solving a potentially difficult minimum-norm constrained optimization problem at each trading time $t$ and in each state $(x, G_n(t))$, FTFH aims heuristically to take small futures trading positions. Based on Proposition 1, if in (9) we express the inverse of the matrix $B_n^{-1}$ as $\zeta_n^{-1}/|B_n^{-1}|$, where $\zeta_n^{-1}$ is the matrix of cofactors of $B_n^{-1}$ (if $K = 1$ then the cofactor matrix is just the scalar 1), then we can rewrite (9) as

$$q^{\pi, H_n}(t) = \Delta_n^{\pi, H_n}(t) + \text{diag}^{-1}(G^{H_n}(t)) \frac{\zeta_n^{-1}}{|B_n^{-1}|} E_{n-1}^{T} \text{diag}(G_{\pi}(t)) \Delta_n^{\pi, H_n}(t).$$

(11)

FTFH selects the set of $K$ contracts to trade that has the largest determinant $|B_n^{-1}|$. This method is computationally efficient because each matrix $B_n^{-1}$, which we refer to as position generating matrix, is chosen once up front (time 0) since this choice only depends on the monthly factor loadings.

7.2 Relationship between BH and MVH
Both BH and FH rely on replication (see §4.1). However, in the presence of term structure model error one may instead focus on finding a minimum variance hedge. We show that BH is an optimal minimum variance hedge. This result thus suggests that insisting on replication based on BH can be useful even in the presence of term structure model error.

The date $t \in [T_n-1, T_n]$ true variance of the sum of the change in the true value of storage $U^{\pi, \diamond}(\cdot)$ over the time interval $[t, t + \Delta t]$ and the time $t + \Delta t$ cash flow from shorting at time $t$ the futures positions in the column vector $q(t)$ is

$$\text{VAR}_{\pi}(U^{\pi, \diamond}(t + \Delta t) - U^{\pi, \diamond}(t) - (q(t))^{T}(G^{\pi}(t + \Delta t) - G^{\pi}(t))),$$

(12)

where $G^{\pi}(\cdot)$ is the true forward curve. Assuming a hypothetically true futures curve evolution model of the type discussed in §6 and letting its deltas be $\Delta_n^{\pi, \diamond}(t)$, a first-order Taylor series approximation for the random change in the true storage value over $[t, t + \Delta t]$ is $(\Delta_n^{\pi, \diamond}(t))^{T}(G^{\pi}(t + \Delta t) - G^{\pi}(t))$. Given that $G^{\pi}(t)$ is known at date $t$, the variance in (12) can be approximated
as \( [q(t) - \Delta^\pi_n(t)]^\top \text{COV}_t^\pi(G^\pi(t + \Delta t)) [q(t) - \Delta^\pi_n(t)] \), where \( \text{COV}_t^\pi(G^\pi(t + \Delta t)) \) is the true covariance matrix of \( G^\pi(t + \Delta t) \) conditional on \( G^\pi(t) \). The resulting MVH model is

\[
\min_{q(t)} [q(t) - \Delta^\pi_n(t)]^\top \text{COV}_t^\pi(G^\pi(t + \Delta t)) [q(t) - \Delta^\pi_n(t)].
\]

(13)

There are two difficulties with this model: Both the deltas \( \Delta^\pi_n(t) \) and the covariance matrix \( \text{COV}_t^\pi(G^\pi(t + \Delta t)) \) are unknown. Despite the second difficulty, Proposition 4 shows that BH is an optimal MVH method when the unknown deltas, \( \Delta^\pi_n(t) \), are replaced with the known deltas obtained under a given \( K \) common factor model, \( \Delta^\pi_n(t) \).

**Proposition 4** (BH optimality). BH optimally solves the version of optimization model (13) obtained by replacing \( \Delta^\pi_n(t) \) with \( \Delta^\pi_n(t) \).

### 7.3 CMVH

With long-tenor storage contracts, the missing contracts friction renders BH impractical and also invalidates FTFH if some of the futures contracts chosen by this method are long-dated. We thus explore a constrained version of the minimum variance model (13) that simply finds the minimum variance hedge given a set of liquid futures contracts. We model the missing contracts friction by supposing that on date \( t \in [T_{n-1}, T_n) \) only the first \( L < N - n \) futures contracts are tradable and, hence, the positions in all of the remaining \( N - n - L \) contracts must be zero. We let \( Q^L_n \) denote the set of all futures position vectors that satisfy this condition.

As will become evident below, the constrained minimum variance hedge depends on the chosen covariance matrix. Although other choices are possible, we proceed by using the covariance matrix \( \text{COV}_t^{N-1}(G(t + \Delta t)) \) of the estimated full dimensional specification of price model (2), i.e., the specification of this price model with \( N - 1 \) common factors on date \( t \). In particular, the entry in position \((j, m)\) of \( \text{COV}_t^{N-1}(G(t + \Delta t)) \) is

\[
F(t, T_j)F(t, T_m) \left[ \exp \left( \Delta t \sum_{k=1}^{N-1} \sigma_{j,k,n-1} \sigma_{m,k,n-1} \right) - 1 \right].
\]

(14)

The CMVH model is

\[
\min_{q(t) \in Q^L_n} [q(t) - \Delta^\pi_n(t)]^\top \text{COV}_t^{N-1}(G(t + \Delta t)) [q(t) - \Delta^\pi_n(t)].
\]

(15)

Proposition 5 characterizes the optimal solution to model (15). Define \( \xi_{n-1}^{N-1} \) as the matrix that includes in position \((j, m)\) the term in square brackets in (14). We include in set \( L_n \) the labels
of the maturities of the $L$ shortest maturity futures that can be used for hedging, i.e., the liquid futures, and in $\mathcal{Z}_n$ the labels of the remaining futures. The matrices $\xi_{n-1}^{N-1,\mathcal{L}_n}$ and $\xi_{n-1}^{N-1,\mathcal{Z}_n}$ are the first $L$ rows and $L$ columns and the first $L$ rows and the last $N - n - L$ columns of the matrix $\xi_{n-1}^{N-1}$. The vectors $G_{\mathcal{L}_n}(t)$ and $G_{\mathcal{Z}_n}(t)$ and $\Delta_{n,\mathcal{L}_n}(t)$ and $\Delta_{n,\mathcal{Z}_n}(t)$ include the prices and deltas corresponding to futures with maturity labels in sets $\mathcal{L}_n$ and $\mathcal{Z}_n$. The vector $q_{\pi,\mathcal{L}_n}(t)$ contains the CMVH trading positions in futures with maturity labels in set $\mathcal{L}_n$.

**Proposition 5** (CMVH positions). If the matrix $\xi_{n-1}^{N-1,\mathcal{L}_n}$ is positive definite, the CMVH futures positions for maturity labels in set $\mathcal{L}_n$ are

$$q_{\pi,\mathcal{L}_n}(t) = \Delta_{n,\mathcal{L}_n}(t) + \text{diag}^{-1}(G_{\mathcal{L}_n}(t)) \left( \xi_{n-1}^{N-1,\mathcal{L}_n} \right)^{-1} \xi_{n-1}^{N-1,\mathcal{Z}_n} \text{diag}(G_{\mathcal{Z}_n}(t)) \Delta_{n,\mathcal{Z}_n}(t).$$ (16)

The CMVH hedge in (16) resembles the hedge in (9) for NFH when CMVH and NFH trade the same futures contracts. In general, however, the two solutions are different. Proposition 6 shows that these solutions are approximately equal in a specific case. Let $\text{COV}_t(G(t + \Delta t))$ be the analogue of $\text{COV}_t^{N-1}(G(t + \Delta t))$ for a $K$ factor model. Recall that $B_{n-1}$ is the position generating matrix corresponding to the $K$ shortest maturity contracts, which are used by NFH.

**Proposition 6** (Approximate equivalence). Suppose that the set $\mathcal{L}_n$ of maturity labels for the tradable contracts includes the labels of the $K$ shortest maturity contracts ($L = K$), the matrix $\text{COV}_t^{N-1}(G(t + \Delta t))$ in (15) is replaced by the matrix $\text{COV}_t(G(t + \Delta t))$, and the position generating matrix $B_{n-1}$ is invertible. If in this version of model (15) each element of the matrix $\text{COV}_t(G(t + \Delta t))$ is approximated by its first-order Taylor expansion around $\Delta t$ equal to 0, then the resulting constrained minimum variance hedge is equal to the naive factor hedge.

We expect the CMVH hedge in (16) to perform better than the NFH hedge in the presence of omitted factors or noise. In this situation, the positions taken by NFH can be extreme, as previously mentioned, because this method attempts to fully eliminate assumed variability while being oblivious of omitted variability in the true model. In contrast, when the rank of the assumed covariance matrix is greater than the number of tradable futures contracts, our implementation of CMVH always takes into account the impact of futures positions on some amount of unavoidable residual futures randomness, which will tend to keep futures position sizes from being too large.

### 8. Numerical Analysis

In this section we quantify the impact of model error on storage valuation (§8.1) and hedging (§8.2), also assessing the performance of the model error mitigation hedging methods proposed in §7.
Figure 5: The value of storage under different term structure models on different valuation dates (the horizontal lines correspond to the hypothetically true models in our model error template in Figure 4).

8.1 Valuation Results
The specific gas storage contracts we consider are related to the twenty-four month gas storage valuation analysis of Lai et al. (2010). We normalize the maximum storage space, \( x \), to 1 mmBtu and set the initial inventory, \( x_0 \), equal to 0, the injection and withdrawal capacities, \( c^I \) and \( c^W \), respectively, equal to \(-0.3\) and \(0.6\) mmBtu per month (i.e., 3.33 months to fill up and 1.67 months to empty), the injection and withdrawal marginal costs, \( c^I \) and \( c^W \), respectively, equal to $0.02/mmBtu and $0.01/mmBtu, and the injection and withdrawal fuel losses, \( \phi^I \) and \( \phi^W \), respectively, equal to 1.01 and 0.99. These parameters are realistic for natural gas storage contracts.

We first investigate the valuation of storage using the RI policy, \( V_0^{RI}(0, F_0) \), given different term structure models. We consider four valuation dates corresponding to the first trading day on different seasons of the year – 3/1/2012 (Spring), 6/1/2012 (Summer), 9/4/2012 (Fall), and
Table 1: The impact of model error on storage valuation: Ratios of storage values obtained with different assumed factor models and different hypothetical true term structure models (Spring, Summer, Fall, and Winter are abbreviated to Sp, Su, Fa, and Wi, respectively).

<table>
<thead>
<tr>
<th>$K$</th>
<th>DGNF = 3</th>
<th>DGNF = 23</th>
<th>DGNF = 3 plus Noise</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Sp</td>
<td>Su</td>
<td>Fa</td>
</tr>
<tr>
<td>1</td>
<td>0.88</td>
<td>0.85</td>
<td>0.80</td>
</tr>
<tr>
<td>2</td>
<td>0.96</td>
<td>0.94</td>
<td>0.93</td>
</tr>
<tr>
<td>3</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>5</td>
<td>1.03</td>
<td>1.03</td>
<td>1.04</td>
</tr>
<tr>
<td>10</td>
<td>1.06</td>
<td>1.06</td>
<td>1.08</td>
</tr>
<tr>
<td>20</td>
<td>1.06</td>
<td>1.05</td>
<td>1.08</td>
</tr>
<tr>
<td>23</td>
<td>1.06</td>
<td>1.05</td>
<td>1.08</td>
</tr>
</tbody>
</table>

12/3/2012 (Winter) – with annualized short-term risk-free Treasury rates equal, respectively, to 0.18%, 0.17%, 0.16%, and 0.18%. Spring and Winter are part of the heating season. Summer and Fall are not. We assume that the storage contract starts on the valuation date (i.e., there are $N = 24$ dates with cash flows). Figure 5 shows the range of possible storage valuations using different term structure models. The Monte Carlo standard errors for these valuations are between $0.013 and $0.016. Compared to the other seasons, storage values starting in Winter are lower because Winter is a high-price season when inventory is typically sold, but here the initial inventory is assumed to be zero. As expected given our PCA results, the value of storage based on common factor models initially increases with the number of factors but largely levels off once the most important statistical factors are included. Also as expected, the storage valuations based on the family of common factor plus noise models are (i) initially higher than the common factor model valuations with the same number of common factors (since there is more transitory price variability) and (ii) decreasing as the most important common factors are added (and idiosyncratic variability is replaced with correlated variability) and eventually largely flattening out.

Using our model error template (Figure 4), Table 1 displays the ratios of the storage valuations corresponding to a subset of the different assumed numbers of common factors, $K$, relative to the valuations of the three data generating models. A number larger (smaller) than 1 means that the assumed model overvalues (undervalues) storage relative to the hypothetically correct valuation. As expected, omitting large common factors (e.g., $K = 1$ and DGNF = 3) leads to substantial undervaluation. More surprisingly, however, the model errors associated with the last 1-2% of the empirical variability cause the storage valuations to change by roughly between −14% and +14% depending on whether that empirical variability is due to missing noise ($K = 3$ and DGNF = 23...
plus noise in Winter) or spurious common factors ($K = 23$ and DGNF = 3 in Winter). In contrast, model error due to missing small common factors has a somewhat lesser impact on storage valuation ($K \geq 5$ and DGNF = 23 across seasons). However, these small common factors effectively add noise to a factor model ($K = 23$ and DGNF = 23 plus noise across seasons). In other words, adding spurious common factors to a factor model can be interpreted as a way of adding some noise to this model.

![Graphs of Flow Rate, Average Inventory, and Flow Time](image)

Figure 6: The operational effect on the value of storage of increasing the number of factors in price model (2).

We conclude this analysis by providing an operational explanation of the largely increasing behavior of the value of storage in Figure 5 and Table 1 as the number of assumed factors in price model (2) increases. Specifically, for each assumed exact factor model we compute the average amount of natural gas sold per stage (i.e., the flow rate) across all the twenty-four stages and the states encountered during the Monte Carlo simulation used for valuation as well as the average inventory across these stages and states. The ratio of average inventory to flow rate is the average number of months that a given unit of purchased natural gas spends in inventory before being sold back to the market (i.e., the flow time from Little’s law). Panels (a)-(c) of Figure 6 show the estimated flow rate, average inventory, and flow time. The flow rate and the average inventory increase and decrease, respectively, and thus the average flow time shortens when increasing the number of factors, $K$ – the increase in the flow rate is due to the increased transitory price variability to trade when adding factors. These patterns taper off at about the same number of factors when the storage value levels off in Figure 5, due to the declining incremental magnitude of this transitory variability. The consequent increased storage value is brought about by trading more natural gas, holding less inventory, and, hence, increasing the frequency of trading.
8.2 Hedging Results
We discuss how we measure hedging effectiveness in §8.2.1. We present our results based on simulated and historical data in §8.2.2 and §8.2.3, respectively.

8.2.1 Measuring Hedging Effectiveness
As natural gas storage is not traded in a liquid market, market prices for natural gas storage are not available to measure hedging effectiveness. Thus, we focus on how effective futures hedging is in reducing the variance of the cash flows generated by physical gas trading using the RI policy along a futures curve sample path. Given an initial futures curve at time 0, let \( \omega \) denote a sample path of \( N \) futures curve realizations at times 0 through \( T_{N-1} \). In our analysis, we use futures price paths both from simulation according to our model error template (Figure 4) and from historical data.

To simplify bookkeeping, we start the storage contract in stage 1 with zero inventory and value it at time 0, so that the initial value of storage is \( U^\text{RI}_1(0, 0, F'_0) \). Thus, there are operational (physical) cash flows on twenty-three monthly stages. We denote by \( P(\omega) \) the time 0 discounted value of the cumulative operating cash flows from time \( T_1 \) through time \( T_{N-1} \) along path \( \omega: \sum_{n \in \mathbb{N} \setminus \{0\}} \delta^n p(a_n^\text{RI}(x_n, F_n), s_n; \omega) \), where \( a_n^\text{RI}(x_n, F_n) \) is the action taken by the RI policy in state \( (x_n, F_n) \) at stage \( n \). We denote by \( \psi_n \left( q_n^\text{RI}(T_{n-1}); \omega \right) \) the cash flow generated at time \( T_n \) along path \( \omega \) by the futures trading positions \( q_n^\text{RI}(T_{n-1}) \): \( \sum_{m \in \mathbb{N}_n} q_{n,m}^\text{RI}(T_{n-1})(F_{n,m} - F_{n-1,m})(\omega) \). We let \( \Psi(\omega) \) be the time 0 discounted value of these trading cash flows from time \( T_1 \) through time \( T_{N-1} \): \( \sum_{n \in \mathbb{N} \setminus \{0\}} \delta^n \psi_n \left( q_{n-1}^\text{RI}(T_{n-1}); \omega \right) \). Our analysis is for monthly futures rebalancing, but more frequent rebalancing just involves increasing the simulation granularity.

From a merchant’s perspective, the reduction in physical cash flow variance due to hedging is a natural metric of hedging effectiveness. We denote by \( \text{VAR}^\circ(P| x_1, F'_0) \) and \( \text{VAR}^\circ(P - \Psi| x_1, F'_0) \) the cross-path variances of the total physical cash flows, \( P \), and residual total cash flows net of hedging, \( P - \Psi \), respectively, given \( x_1 \) and \( F'_0 \), under the model that generates the price data. Our hedging effectiveness metric (HEM) is the reduction in \( \text{VAR}^\circ(P| x_1, F'_0) \) due to hedging:

\[
\text{HEM} := 100 \cdot \left[ 1 - \frac{\text{VAR}(P - \Psi| x_1, F'_0)}{\text{VAR}(P| x_1, F'_0)} \right].
\]

For each pair of an assumed factor model and a hedging method that we consider, we use Monte Carlo simulation under a hypothetical true futures price model to estimate \( \text{VAR}^\circ(P| x_1, F'_0) \) and \( \text{VAR}^\circ(P - \Psi| x_1, F'_0) \), and hence HEM. HEM equals 100 with a perfect hedge. In this case, \( P(\omega) - \Psi(\omega) \) is equal to \( U^\text{RI}_1(0, 0, F'_0) \), a constant given \( x_1 \) (= 0) and \( F'_0 \), for almost every sample path \( \omega \).

Empirical backtesting of hedging using historical data involves following a hedging strategy over a series of subintervals, each of which represents one observation in the sample. A complication
is that the initial futures curve, $F'_0$, and hence the initial valuation, $U_{RI}^1(0, 0, F'_0)$, vary across subintervals (i.e., observations). We thus modify HEM using the sample variances of $P - \Psi - U_{RI}^1(0, 0, F'_0)$ and $P - U_{RI}^1(0, 0, F'_0)$ to control for the different initial values across observations. We denote by BHEM the resulting backtesting version of HEM.

**8.2.2 Hedging Results with Simulated Data**

Given paths of simulated prices from each one of the hypothetically true data generating models in our model error template (Figure 4), we implement and evaluate different hedging strategies based on a variety of assumed common factor models and their associated deltas. The PCA loading coefficients of factors shared by the assumed and hypothetically true models are taken to be the same (i.e., there is no parameter error). The resulting HEM estimates are then compared to assess the sensitivities of different hedging strategies to model error.

For brevity, we focus on the Summer instance and consider a subset of the possible values of the number of assumed common factors, $K$. We take the NYMEX closing futures prices on 6/1/2012 as the initial time 0 futures curve. The financial trading positions are rebalanced once every month. We use a Monte Carlo sample size of 200 paths to estimate HEM for each assumed-model/hypothetically-true-model/hedge-strategy triple that we consider. Rebalancing requires nested simulations to compute deltas, for which we use 10,000 additional Monte Carlo samples. Obtaining the cash flows associated with a given hedging strategy along one sample path is computationally intensive, requiring about 7 Cpu minutes, depending on the price models and hedging method employed – we use the gcc version 4.8.2 2013.10.17 (Red Hat 4.8.2-1) compiler on a 64 bits PowerEdge R515 with twelve AMD Opteron 4176 2.4GHz processors, each with 64 GB of memory running Linux Fedora 19 (the stated Cpu times correspond to using a single processor). We use common random numbers when generating price samples across different models that share common factors.

Table 2 reports the HEM estimates for NFH, FTFH, and BH – the standard errors of these estimates (computed by bootstrapping) are in Online Appendix C. These hedging strategies are all near optimal (HEM close to 100) in the absence of model error (DGNF = $K = 3$ or 23; in the latter case these strategies coincide). These HEM estimates are not equal to 100 due to simulation, discrete rebalancing, and delta computation errors. The precisions of these HEM estimates are comparable. With spurious factors in the assumed model (DGNF = $K < 3$), NFH, FTFH, and BH all perform well (estimated HEM $\geq 98.00$), with one exception (estimated HEM = 73.25 with $K = 5$ for NFH). This good performance is expected, since overhedging against additional zero-probability futures curve changes has no impact on the effectiveness of hedges against assumed futures curve
Table 2: The impact of model error on hedging with simulated data: HEM estimates for NFH, FTFH, and BH

<table>
<thead>
<tr>
<th>$K$</th>
<th>DGNF = 3</th>
<th></th>
<th>DGNF = 23</th>
<th></th>
<th>DGNF = 3 plus Noise</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>NFH</td>
<td>FTFH</td>
<td>BH</td>
<td>NFH</td>
<td>FTFH</td>
</tr>
<tr>
<td>1</td>
<td>87.79</td>
<td>87.40</td>
<td>98.14</td>
<td>87.73</td>
<td>87.33</td>
</tr>
<tr>
<td>2</td>
<td>49.96</td>
<td>96.56</td>
<td>98.80</td>
<td>−246.73</td>
<td>96.51</td>
</tr>
<tr>
<td>3</td>
<td>97.55</td>
<td>98.87</td>
<td>98.87</td>
<td>−1011.47</td>
<td>98.43</td>
</tr>
<tr>
<td>5</td>
<td>73.25</td>
<td>98.82</td>
<td>98.80</td>
<td>−5838.15</td>
<td>98.88</td>
</tr>
<tr>
<td>10</td>
<td>98.00</td>
<td>98.80</td>
<td>98.77</td>
<td>−2522.85</td>
<td>99.08</td>
</tr>
<tr>
<td>15</td>
<td>98.30</td>
<td>98.79</td>
<td>98.76</td>
<td>−1885.37</td>
<td>99.07</td>
</tr>
<tr>
<td>20</td>
<td>98.69</td>
<td>98.78</td>
<td>98.76</td>
<td>−227.40</td>
<td>99.08</td>
</tr>
<tr>
<td>23</td>
<td>98.76</td>
<td>98.78</td>
<td>98.76</td>
<td>99.08</td>
<td>99.08</td>
</tr>
</tbody>
</table>

changes that actually can occur in the hypothetically true process. All the corresponding HEM estimates have similar precisions. The exception for NFH is due to the amplification by the large trading positions of NFH, discussed below, of delta estimation and discrete rebalancing errors, which also makes the resulting HEM estimate much less precise than the other HEM estimates (for the spurious factor case). The performance of different hedging strategies diverges sharply in the presence of missing factors or missing noise.

**NFH.** With omitted important factors (DGNF = 3 > $K$), NFH with $K = 1$ assumed factor performs reasonably well (estimated HEM $\geq 87.79$) but its performance actually degrades with $K = 2$ assumed factors (estimated HEM = 49.96). With more omitted factors (DGNF = 23 > $K$), NFH with $K = 1$ assumed factor exhibits reasonable performance (estimate HEM = 87.73) but with 2 or more assumed factors performs disastrously: The HEM estimates are large and negative. In other words, NFH dramatically increases residual cash flow variance relative to the unhedged cash flow variance. It is striking how large the effect on NFH is, since the omitted factors with $K \geq 3$ only contribute at most 1-2% to the log futures price return variance. The effect on NFH of missing noise, combined with missing the second factor or with spurious factors (DGNF = 3 plus noise) is similarly bad, with the exceptions of the special cases $K = 1$ and 23.

The disastrous performance of NFH is a consequence of this strategy taking extremely large trading positions that magnify small model errors (and make the resulting HEM estimates less precise than in the case of smaller trading positions). Table 3 displays the average cumulative long and short futures positions taken by NFH along our simulated price paths for different hypothetical true models. The positions with $K > 1$ are very large. These large position sizes are induced by the small determinants $|B_{n-1}|$ of the position generating matrices in the denominator of the second term on the right-hand side of expression (11), which is, in turn, a consequence of the fact that
Table 3: Average cumulative futures positions taken by NFH.

<table>
<thead>
<tr>
<th>$K$</th>
<th>DGNF = 3</th>
<th>DGNF = 23</th>
<th>DGNF = 3 plus Noise</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Long</td>
<td>Short</td>
<td>Long</td>
</tr>
<tr>
<td>1</td>
<td>12.32</td>
<td>0.37</td>
<td>12.33</td>
</tr>
<tr>
<td>2</td>
<td>145.45</td>
<td>135.84</td>
<td>146.87</td>
</tr>
<tr>
<td>3</td>
<td>483.96</td>
<td>464.89</td>
<td>486.25</td>
</tr>
<tr>
<td>5</td>
<td>2332.76</td>
<td>2308.64</td>
<td>2351.07</td>
</tr>
<tr>
<td>10</td>
<td>2683.42</td>
<td>2665.49</td>
<td>2693.93</td>
</tr>
<tr>
<td>15</td>
<td>2336.51</td>
<td>2321.98</td>
<td>2328.71</td>
</tr>
<tr>
<td>20</td>
<td>927.05</td>
<td>908.81</td>
<td>924.48</td>
</tr>
<tr>
<td>23</td>
<td>25.49</td>
<td>8.02</td>
<td>25.84</td>
</tr>
</tbody>
</table>

the factor loadings of short-dated futures prices are rather similar. With no model error, these positions, even if very large (DGNF = $K = 3$), hedge the value of storage well because they are consistent with price changes from the data generating model (although they might be impractical to trade when large). With underhedging due to omitted factors or noise, however, these positions are structurally inconsistent with the realized price changes, and their size magnifies the omitted randomness. This inconsistency generates severe discrepancies between the realized physical trading cash flows and the hedging cash flows. The good performance of NFH when $K$ is equal to 1 and 23 is explained by the matrix inversion in (9), which is equivalent to (11), reducing to a division by a scalar, which does not generate huge trading positions, when $K = 1$, and by NFH reducing to BH, which is not based on (9), when $K = 23$.

**FTFH.** The poor performance of NFH motivated us to develop a hedging strategy with small futures positions; FTFH heuristically achieves this goal (see §7.1). Although we expect FTFH to perform better than NFH with underhedging due to omitted factors or noise, how much was unclear a priori. Our results indicate that the improvement is dramatic. Specifically, the estimated HEMs of FTFH in Table 2 are 96.51 or higher and dominate the ones for NFH in all cases irrespective of the type of model error whenever $K > 1$ (the two strategies are identical with $K = 1$ when DGNF=23 or DGNF=3 plus noise). In fact, FTFH achieves near optimal performance (i.e., HEM estimates close to 100 and with high precisions) when the number of assumed factors $K$ is sufficiently large. The improvement in performance of FTFH relative to NFH is due to the smaller trading position sizes of FTFH: The average cumulative long (short) trading positions vary from 12.41 to 28.41 (0.35 to 9.97) across the considered cases. This finding underscores the impact of carefully selecting the particular traded contracts when implementing FH.

**BH.** The HEM estimates for BH in Table 2 vary between 98.14 and 99.33 and are highly precise. Consistent with the bounds in Proposition 3, the smallest and largest realized BH trading
Table 4: The impact of model errors and missing contracts on hedging with simulated data: HEM estimates for CMVH implemented using the first $L = 3, 5,$ and $10$ front-end futures.

<table>
<thead>
<tr>
<th>$K$</th>
<th>$L = 3$</th>
<th>$L = 5$</th>
<th>$L = 10$</th>
<th>$L = 3$</th>
<th>$L = 5$</th>
<th>$L = 10$</th>
<th>$L = 3$</th>
<th>$L = 5$</th>
<th>$L = 10$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>93.22</td>
<td>95.26</td>
<td>98.11</td>
<td>90.80</td>
<td>93.74</td>
<td>97.33</td>
<td>91.00</td>
<td>93.73</td>
<td>96.74</td>
</tr>
<tr>
<td>2</td>
<td>93.74</td>
<td>95.92</td>
<td>98.63</td>
<td>91.65</td>
<td>94.50</td>
<td>97.91</td>
<td>91.91</td>
<td>94.55</td>
<td>97.56</td>
</tr>
<tr>
<td>3</td>
<td>93.89</td>
<td>96.05</td>
<td>98.67</td>
<td>91.92</td>
<td>94.64</td>
<td>97.94</td>
<td>92.29</td>
<td>94.78</td>
<td>97.80</td>
</tr>
<tr>
<td>5</td>
<td>93.94</td>
<td>96.09</td>
<td>98.64</td>
<td>92.07</td>
<td>94.73</td>
<td>97.93</td>
<td>92.48</td>
<td>94.87</td>
<td>97.86</td>
</tr>
<tr>
<td>10</td>
<td>93.94</td>
<td>96.08</td>
<td>98.62</td>
<td>92.09</td>
<td>94.72</td>
<td>97.90</td>
<td>92.55</td>
<td>94.89</td>
<td>97.87</td>
</tr>
<tr>
<td>15</td>
<td>93.94</td>
<td>96.08</td>
<td>98.62</td>
<td>92.08</td>
<td>94.71</td>
<td>97.89</td>
<td>92.57</td>
<td>94.90</td>
<td>97.88</td>
</tr>
<tr>
<td>20</td>
<td>93.94</td>
<td>96.08</td>
<td>98.62</td>
<td>92.09</td>
<td>94.71</td>
<td>97.89</td>
<td>92.57</td>
<td>94.89</td>
<td>97.89</td>
</tr>
<tr>
<td>23</td>
<td>93.94</td>
<td>96.08</td>
<td>98.62</td>
<td>92.09</td>
<td>94.71</td>
<td>97.89</td>
<td>92.57</td>
<td>94.89</td>
<td>97.88</td>
</tr>
</tbody>
</table>

positions (i.e., deltas) across all sample paths, trading dates, and contracts traded are $-0.30$ and $0.59$ irrespective of the assumed number of factors, $K$, used to compute the deltas and the hypothetically true model that generates the data. The average cumulative long and short positions vary in between 24.90 and 27.48 and 8.02 and 10.98, respectively. BH outperforms FTFH. The similar performance of BH across these different cases is remarkable.

The difference in performance between NFH and BH can be explained by the fact that FH attempts to hedge common randomness across futures prices by relying on specific assumptions about the structure of that randomness via (9). If actual futures price dynamics include additional or different types of randomness that the assumed model omits, this discrepancy can cause hedging errors. In contrast, given the option deltas, BH does not rely on the specific structure of the futures curve dynamics, i.e., it is not based on (9); instead, each dated cash flow is hedged individually using futures contracts with the matching maturity. Consequently, BH is an effective approach to reduce the impact of model error on hedging. This numerical finding is consistent with Proposition 4, which shows that BH is optimal in a minimum variance sense even if the true futures curve covariance matrix is unknown.

**CMVH.** We investigate the impact of the missing contracts friction on hedging by analyzing the performance of CMVH implemented when trading is limited to 3, 5, and 10 front-end futures contracts ($L = 3, 5,$ and $10$). Table 4 shows that the performance of CMVH is good – the standard errors of the HEM estimates in this table (computed via bootstrapping) are available in Online Appendix C. The improvement using 5 versus 3 contracts is appreciable, and the improvement from using 10 versus 5 contracts is still non-trivial. The performance of CMVH is inferior to the one of BH, because, as shown in Proposition 4, BH is the optimal *unconstrained* MVH approach.
Table 5: The impact of model error on hedging with real data: BHEM estimates for NFH, FTFH, and BH.

<table>
<thead>
<tr>
<th>$K$</th>
<th>In-Sample</th>
<th>Out-of-Sample</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$K$</td>
<td>NFH</td>
</tr>
<tr>
<td>1</td>
<td>79.90</td>
<td>99.33</td>
</tr>
<tr>
<td>2</td>
<td>-2300.25</td>
<td>99.00</td>
</tr>
<tr>
<td>3</td>
<td>-6726.48</td>
<td>99.38</td>
</tr>
<tr>
<td>5</td>
<td>-1302.06</td>
<td>99.12</td>
</tr>
<tr>
<td>10</td>
<td>-1327.40</td>
<td>99.33</td>
</tr>
<tr>
<td>15</td>
<td>14.59</td>
<td>99.41</td>
</tr>
<tr>
<td>20</td>
<td>97.93</td>
<td>99.40</td>
</tr>
<tr>
<td>23</td>
<td>99.39</td>
<td>99.39</td>
</tr>
</tbody>
</table>

However, CMVH can still be near optimal in most cases when $L = 10$. The precisions of the HEM estimates in Table 4 increase with the number of traded contracts, $L$, given the number of assumed common factors, $K$, and the hypothetically true data generating model. The CMVH trading positions are not large, despite being limited to just the front-end contracts: The average cumulative long (short) trading positions ranges for $L$ equal to 3, 5, and 10, respectively, are 20.70-22.09 (8.71-9.74), 25.59-27.22 (11.87-13.58), and 30.17-32.00 (13.43-16.48). Moreover, CMVH uses the same contracts as NFH when $K = L = 3, 5, and 10$. Thus, the good CMVH performance highlights the impact not just of which contacts to trade (as with FTFH) but also of the specific positions themselves as compared to NFH.

8.2.3 Hedging Results with Historical Data

We conclude with backtested hedging results for actual historical futures price data. Since the true price model is unknown, our backtesting analysis combines all of the possible model errors, including an unknown number of factors, parameter estimation errors, and misspecification due to possible stochastic volatility and jumps. Because history contains a single path, we split the data into rolling two-year subperiods to obtain multiple subpaths. We perform 15 in-sample experiments (for subperiods 1997-1998, 1998-1999, . . . , 2011-2012) and 13 out-of-sample experiments (for 1999-2000 through 2011-2012) where the model is estimated over the preceding two years.

The hedging performance results using actual data in Table 5 largely confirm our results with simulated data (standard errors are in Online Appendix C): NFH can perform disastrously (except when $K = 1, 20, and 23$, where the last case corresponds to BH), FTFH is effective once the most important statistical factors are included, and BH is still more effective. The BHEM estimates for FTFH and BH are more precise than the BHEM estimates for NFH (except for $K = 1$ when comparing FTFH and NFH). With a few exceptions, the BHEM estimates are less precise in the out-of-sample case than in the in-sample case. Table 6, with standard errors in Online Appendix C,
Table 6: The impact of model errors and missing contracts on hedging with real data: BHEM estimates for CMVH.

<table>
<thead>
<tr>
<th></th>
<th>In-Sample</th>
<th></th>
<th>Out-of-Sample</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>L</td>
<td>3</td>
<td>5</td>
<td>10</td>
</tr>
<tr>
<td>K</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>82.60</td>
<td>91.12</td>
<td>98.27</td>
<td>81.60</td>
</tr>
<tr>
<td>2</td>
<td>84.69</td>
<td>92.36</td>
<td>98.44</td>
<td>85.40</td>
</tr>
<tr>
<td>3</td>
<td>87.29</td>
<td>94.06</td>
<td>98.83</td>
<td>86.33</td>
</tr>
<tr>
<td>5</td>
<td>88.26</td>
<td>94.63</td>
<td>98.91</td>
<td>86.62</td>
</tr>
<tr>
<td>10</td>
<td>88.81</td>
<td>95.03</td>
<td>99.00</td>
<td>86.82</td>
</tr>
<tr>
<td>15</td>
<td>88.83</td>
<td>95.04</td>
<td>99.00</td>
<td>86.82</td>
</tr>
<tr>
<td>20</td>
<td>88.83</td>
<td>95.04</td>
<td>99.00</td>
<td>86.84</td>
</tr>
<tr>
<td>23</td>
<td>88.84</td>
<td>95.04</td>
<td>99.00</td>
<td>86.82</td>
</tr>
</tbody>
</table>

confirms the robustness of CMVH using actual historical data when there are liquidity restrictions on which contracts can be traded. When only the first 3 contracts are traded, CMVH performs reasonably well. Its performance improves substantially when the first 5 contracts are available, and it becomes near optimal using the first 10 contracts, in which case it is even slightly better than the performance of BH in the out-of-sample case (see Table 5). The precisions of the BHEM estimates improve when the number of futures contracts available to trade increases. Compared to the in-sample case, in the out-of-sample case the BHEM estimates are generally less precise when $L$ is equal to 3 or 5 and more precise when $L$ is equal to 10. Thus, CMVH appears to be an effective method in practice despite the presence of various model errors and the missing contracts friction.

9. Conclusions

In this paper we show that various types of small empirically calibrated term structure model errors substantially impact the valuation and hedging of merchant commodity storage. We propose several near optimal approaches to mitigate the negative effects of these errors on hedging, also considering liquidity constraints. Beyond our application to the merchant storage of natural gas, our research has potential relevance in other merchant operations contexts when valuing and hedging real and financial options that are contingent on futures term structure dynamics, or when deriving inventory/production management and capacity investment policies that depend on demand forecast term structures and for which financial hedging can also be relevant. Further research might focus on considering other hedging approaches, e.g., gamma and vega hedging, investigating the effect of limited liquidity, such as parameter estimation error, on the calibration of futures curve term structure models, and extending our model error analysis to incorporate transaction costs (bid-ask spreads) in physical and financial trading and their impact of operating and hedging policies.
References


EIA. 2011c. Peak underground working natural gas storage capacity.  
http://www.eia.gov/naturalgas/storagecapacity/.

Energy and Power Risk Management: New Developments in Modeling,  

Fan, R., A. Gupta, P. Ritchken. 2001. On pricing and hedging in the swaption market: How many factors, really?  
Working Paper, Case Western Reserve University, Cleveland, OH, USA.

Frestad, D. 2008. Common and unique factors influencing daily swap returns in the Nordic electricity market,  
Energy Economics 30(3) 1081–1097.

Energy Economics 34(5) 1341–1355.

Manufacturing & Service Operations Management 7(2) 103–120.

and Energy. John Wiley & Sons, Chichester, UK.

Management Science 51(7) 1076–1091.

Monte Carlo Methods in Financial Engineering. Springer, New York, NY, USA.

Management Science 57(12) 2228–2244.

S. Axsaeter, C. Schneeweiss, E. Silver, eds., Multi-Stage Production Planning and Inventory Control.  

Commodities Now September 1–5.

Commodities Now June 1–4.

Heath, D., P. L. Jackson. 1994. Modeling the evolution of demand forecasts with application to safety stock  
analysis in production/distribution systems.  
IIE Transactions 26(3) 17–30.


Management Science 50(7) 909–921.

Kennedy, D. P. 1994. The term structure of interest rates as a Gaussian random field.  
Mathematical Finance 4(3) 247–258.

Kouvelis, P., R. Li, Q. Ding. 2013. Managing storable commodity risks: The role of inventory and financial  
hedge.  

practice-based heuristics for natural gas storage valuation.  
Operations Research 58(3) 564–582.

natural gas terminal.  
Operations Research 59(3) 602–616.


Longstaff, F. A., P. Santa-Clara, E. S. Schwartz. 2001. The relative valuation of caps and swaptions: Theory  
and empirical evidence.  
Journal of Finance 56(6) 2067–2109.


Applied Mathematical Finance 9(1) 21–43.

Real Options and Energy Management Using Options Methodology to Enhance Capital Budgeting Decisions.  


Thompson, M. 2012. Natural gas storage valuation, optimization, market and credit risk management. Working Paper, Queen’s School of Business, Queen’s University, Kingston, Ontario, CA.


Online Appendix

A. Deltas with Exact Storage Valuation

Proposition 9 in this appendix extends the pathwise delta characterization established in Proposition 2 to the case of exact storage valuation based on the following stochastic dynamic program (SDP) for exact storage valuation (Lai et al. 2010, §2):

\[
V_n(x_n, F_n) = \max_{a \in A(x_n)} p(a, s_n) + \delta \mathbb{E} \left[ V_{n+1}(x_n - a, F_{n+1}) \middle| F_n \right],
\]

for all \( n \in \mathcal{N} \) and \((x_n, F_n) \in \mathcal{X} \times \mathbb{R}^{N-n}_+\), where \( V_n(x_n, F_n) \) denotes the optimal value function in stage \( n \) and state \((x_n, F_n)\), and \( V_N(x_N, F_N) := 0 \), for all \( x_N \in \mathcal{X} \) (recall that \( F_N \equiv 0 \)).

Proposition 7 characterizes the optimal value function and an optimal operating policy of this SDP. The proof of this result is a simple adaptation of the proofs of Lai et al. (2010, Theorem 1) and Secomandi (2010, Theorem 1).

**Proposition 7** (Concavity and basestock optimality). In every stage \( n \), the function \( V_n(x_n, F_n) \) is concave in \( x_n \) for each given \( F_n \), and the optimal policy for the SDP (17) features two basestock targets, \( b_n(F_n) \), \( \bar{b}_n(F_n) \) ∈ \( \mathcal{X} \), such that \( b_n(F_n) \leq \bar{b}_n(F_n) \) and an optimal decision rule \( A^*(x_n, F_n) \) satisfies

\[
A^*_n(x_n, F_n) = \begin{cases} 
C^I \lor [x_n - b_n(F_n)], & x_n \in [0, b_n(F_n)), \\
0, & x_n \in [b_n(F_n), \bar{b}_n(F_n)], \\
C^W \land [x_n - \bar{b}_n(F_n)], & x_n \in (\bar{b}_n(F_n), x] \end{cases},
\]

Proposition 8, based on Lemma 1, establishes basic properties of the optimal value function and basestock targets, which are used to establish Proposition 9. All of these results are based on Assumption 2.

**Assumption 2** (Capacities and maximum space). The capacity limits \( C^I \) and \( C^W \) as well as the maximal inventory level \( x \) are integer multiples of some positive real number \( Q \).

Lemma 1 is related to Secomandi (2010, Propositions 2 and 3).

**Lemma 1** (Characterization). Under Assumption 2, in every stage \( n \in \mathcal{N} \): (a) The function \( V_n(x, F_n) \) is piecewise linear continuous in inventory \( x \in \mathcal{X} \) with break points in set \( \mathcal{Q} := \{0, Q, 2Q, \ldots, x\} \) for each given futures curve \( F_n \in \mathbb{R}^{N-n}_+ \); (b) There exist optimal basestock levels \( b_n(F_n) \) and \( \bar{b}_n(F_n) \) in set \( \mathcal{Q} \), \( \forall F_n \in \mathbb{R}^{N-n}_+ \).
Proposition 8. Under Assumption 2, in every stage \( n \in N \): (a) For all inventory levels \( x \in \mathcal{X} \) and futures curves \( F_n \in \mathbb{R}_+^N \) it holds that \( V_n(x, F_n) = \max_{a \in \mathcal{A}(x)} p(a, s_n) + \delta \mathbb{E} [V_{n+1}(x_n - a, F_{n+1}) | F_n] \); (b) For each given inventory \( x \in \mathcal{X} \) it holds that the function \( V_n(x, F_n) \) is Lipschitz continuous in the futures curve \( F_n \in \mathbb{R}_+^N \); i.e., there exists \( L_n(x) \in \mathbb{R}_+ \) such that \( |V_n(x, F_n^2) - V_n(x, F_n^1)| \leq L_n(x) \sum_{m=n}^N |F_{n,m}^2 - F_{n,m}^1|, \forall F_n^1, F_n^2 \in \mathbb{R}_+^N \).

In Proposition 9 we indicate by \( x_n^* \) the optimal inventory level in stage \( m \in N \setminus \{0\} \) and by \( \Delta_n, m(t, x_n, G_n(t)) \) the delta under an optimal policy.

Proposition 9 (Pathwise deltas with an optimal policy). Under Assumption 2, for every \( n \in N \setminus \{0\} \) it holds that

\[
\Delta_n, m(t, x_n, G_n(t)) = \frac{\mathcal{L}(t, T_m)}{F(t, T_m)} \mathbb{E} [\phi W_1 \{ x_m^* \in [0, \bar{b}_m(F_m)] \} + \phi W_1 \{ x_m^* \in (\bar{b}_m(F_m), \bar{x}) \}] s_m A_m^*(x_m, F_m) | x_n, G_n(t),
\]

for all \( m \in N_n, x_n \in \mathcal{X}, t \in [T_{n-1}, T_n), \) and \( G_n(t) \in \mathbb{R}_+^{N-n} \).

B. Proofs

Additional Notation. We denote by \( Y \) the vector \((Y_k, k \in K)\) of \( K \) uncorrelated standard normals. Given \( n \) and \( m \in N_{n+1}, t \in [T_n, T_{n+1}), \) and \( t' \in (T_n, T_{n+1}] \) with \( t < t' \), we define the quantity \( \beta_m(t, t', Y) \) as

\[
-\mathcal{L}(t, T_m) \mathbb{E} [\phi W_1 \{ x_m^* \in [0, \bar{b}_m(F_m)] \} + \phi W_1 \{ x_m^* \in (\bar{b}_m(F_m), \bar{x}) \}] s_m A_m^*(x_m, F_m) | x_n, G_n(t)
\]

Under Assumption 2, we can equivalently express \( F(t') \) given \( F(t) \) as \( F(t') \beta(t, t', Y) \) and \( F(T_{n+1}) \) given \( F(t) \) as \( F(t') \beta(t, T_{n+1}, Y) \), where the products are in the componentwise sense. The equalities \( F(t, T_m) = \mathbb{E} [F(t', T_m) | F(t, T_m)] \) and \( \mathbb{E} [F(t', T_m) | F(t, T_m)] = F(t, T_m) \mathbb{E} [\beta_m(t, t', Y)] \) imply that

\[
\mathbb{E} [\beta_m(t, t', Y)] = 1.
\]

Given \( n < N - 1 \), when \( t = T_n \) and \( t' = T_{n+1} \), we abbreviate \( \beta_m(T_n, T_{n+1}, Y) \) to \( \beta_{n,n+1}^m(Y) \) for all \( m \in N_{n+1}, \) and \( \beta(T_n, T_{n+1}, Y) \) to \( \beta_{n,n+1}(Y) \). We use this notation in the proofs of Lemma 2 and Proposition 8.
We define by \( W_n^\pi(x, F'_n) := \delta \mathbb{E} [V_{n+1}^\pi(x, F_{n+1}) | F'_n] \) the policy \( \pi \) continuation value in stage \( n < N \) given the inventory level \( x \) and the futures curve \( F'_n \). We use this notation in the proof of Proposition 2. This notation allows us to write

\[
V_n^\pi(x, F_n) = \sum_{m=n}^{N-1} \delta^{m-n} \mathbb{E}[p(A_m^\pi(x_m, F_m), s_m) | x, F_n]
\]

\[
= p(A_n^\pi(x, F_n), s_n) + \sum_{m=n+1}^{N-1} \delta^{m-n} \mathbb{E}[p(A_m^\pi(x_m, F_m), s_m) | x - A_n^\pi(x, F_n), F'_n]
\]

\[
= p(A_n^\pi(x, F_n), s_n) + \delta \mathbb{E} [V_{n+1}^\pi(x - A_n^\pi(x, F_n), F_{n+1}) | F'_n]
\]

\[
= p(A_n^\pi(x, F_n), s_n) + W_n^\pi(x - A_n^\pi(x, F_n), F'_n).
\]

We define by \( W_n(x, F'_n) := \delta \mathbb{E} [V_{n+1}(x, F_{n+1}) | F'_n] \) the optimal continuation value in stage \( n < N \) given the inventory level \( x \) and the futures curve \( F'_n \). We use this notation in the proofs of Lemma 1 and Propositions 8 and 9. We denote by \( v_n(x, a, F_n) \) the objective function of the maximization on the right hand side of the SDP (17). We use this notation in the proofs of Propositions 8 and 9 and Lemma 3.

**Proof of Proposition 1 (FH positions).** The system of linear equations (8) can be expressed as

\[
G^{H_n}(t) (B_{n-1}^T) q^{H_n}(t) = G^{H_n}(t) (B_{n-1}^T) \Delta_n^{\pi,K,H_n} + E_{n-1}^T \text{diag}(G^{H_n}(t)) \Delta_n^{\pi,K,H_n}.
\]

Premultiplying both sides of this expression by \( \text{diag}^{-1}(G^{H_n}(t)) (B_{n-1}^T)^{-1} \) yields the claimed result. \( \square \)

**Lemma 2.** Lemma 2 is needed in the proof of Proposition 2.

**Lemma 2 (Pathwise derivatives).** Under Assumption 1(a), given \( x \in \mathcal{X} \), for every \( n \in \mathcal{N} \setminus \{N-1\} \) and \( m \in \mathcal{N}_{n+1} \) it holds that

\[
\frac{dV_{n+1}^\pi(x, F_{n+1})}{df_{n,m}} = \frac{\partial V_{n+1}^\pi(x, F_{n+1})}{\partial f_{n+1,m}} \frac{f_{n+1,m}}{f_{n,m}}, \tag{21}
\]

\[
\frac{\partial \mathbb{E}[V_{n+1}^\pi(x, F_{n+1}) | F'_n]}{\partial f_{n,m}} = \mathbb{E} \left[ \frac{dV_{n+1}^\pi(x, F_{n+1})}{df_{n,m}} | F'_n \right]. \tag{22}
\]

**Proof.** We interpret \( f_{n,m} \) as a parameter and write \( F_{n+1}(f_{n,m}) \) to explicitly indicate the dependence of \( F_{n+1} \) on this parameter. In particular, as \( F_{n+1,m} = F_{n,m} \beta_m^{n+1}(Y) \), \( F_{n+1,m} \) depends on \( f_{n,m} \) but every other \( f_{n+1,j} \), with \( j > m \), does not.

We first show that conditions (A1)-(A4) in Appendix A of Broadie and Glasserman (1996) hold with respect to \( \mathbb{E}[V_{n+1}^\pi(x, F_{n+1}(f_{n,m})) | F'_n] \).

(A1) Under model (2) the quantity \( \partial F_{n+1,j}(f_{n,m})/\partial f_{n,m} \) exists with probability 1 because it is equal to \( \beta_m^{n+1}(Y) \) if \( j = m \) and to 0 when \( j > m \).
(A2) Fix \( x \) and let \( \mathcal{D}^{V}_{n+1} \) denote the set of futures curves \( F_{n+1} \) at which \( V_{n+1}^{\pi}(x, F_{n+1}) \) is differentiable with respect to each element of \( F_{n+1} \). Assumption 1(a) and Rademacher’s theorem imply that \( V_{n+1}^{\pi}(x, F_{n+1}) \) is differentiable almost everywhere on \( \mathbb{R}^{N-n-1} \) with respect to each element of \( F_{n+1} \). Price model (2) implies that \( \mathbb{P}(F_{n+1}(F_{n,m}) \in \mathcal{D}^{V}_{n+1}|F_{n}^{\prime}) = 1 \), where \( \mathbb{P} \) denotes probability, for all \( F_{n,m} \in \mathbb{R}_{+} \).

(A3) This is Assumption 1(a).

(A4) The random variable \( F_{n+1,m}(F_{n,m}) \) is almost surely Lipschitz continuous with integrable modulus \( \beta^{n,n+1}_{m}(Y) \) because \( |F_{n+1,m}(F_{n,m}^{2}) - F_{n+1,m}(F_{n,m}^{1})| = \beta^{n,n+1}_{m}(Y)|F_{n,m}^{2} - F_{n,m}^{1}| \), for all \( F_{n,m}^{1}, F_{n,m}^{2} \in \mathbb{R}_{+} \), and \( \mathbb{E}[\beta^{n,n+1}_{m}(Y)] = 1 \). Every other random variable \( F_{n+1,j}(F_{n,m}) \), with \( m < j \), is almost surely Lipschitz with integrable modulus 0 because each such random variable does not depend on \( F_{n,m} \).

Following Broadie and Glasserman (1996, p. 280), (21) holds because under conditions (A1)-(A2) the pathwise derivative \( dV_{n+1}(x, F_{n+1})/dF_{n,m} \) satisfies

\[
dV_{n+1}(x, F_{n+1})/dF_{n,m} = \sum_{j=n+1}^{N-1} \frac{\partial V_{n+1}(x, F_{n+1})}{\partial F_{n+1,j}} \frac{\partial F_{n+1,j}}{\partial F_{n,m}} = \frac{\partial V_{n+1}(x, F_{n+1})}{\partial F_{n+1,m}} \beta^{n,n+1}_{m}(Y)
\]

Expression (22) follows from Proposition 1 in Broadie and Glasserman (1996). □

Proof of Proposition 2 (Pathwise deltas). For simplicity of exposition, we prove the claimed result for \( t = 0 \) and \( n = 1 \). The proof for the general case follows similar steps. As \( F_{0}^{\prime} \equiv G_{1}(T_{0}) \), in this proof we use the notation \( F_{0}^{\prime} \) in lieu of the notation \( G_{1}(T_{0}) \).

Suppose \( m = 1 \). Formula (6) and Lemma 2 imply that

\[
\frac{\Delta_{1,1}^{\pi}(0, x_{1}, F_{0}^{\prime})}{\delta(0, T_{1})} = \frac{1}{\delta(0, T_{1})} \frac{\partial U_{1}^{\pi}(0, x_{1}, F_{0}^{\prime})}{\partial F_{0,1}} = \frac{\partial \mathbb{E}[V_{1}^{\pi}(x_{1}, F_{1})|F_{0}^{\prime}]}{\partial F_{0,1}} = \mathbb{E} \left[ \frac{dV_{1}^{\pi}(x_{1}, F_{1})}{dF_{0,1}} \right]. \tag{23}
\]

It follows from Lemma 2 that

\[
\frac{dV_{1}^{\pi}(x_{1}, F_{1})}{dF_{0,1}} = \sum_{n=2}^{N-1} \frac{\partial V_{1}^{\pi}(x_{1}, F_{1})}{\partial F_{1,m}} \frac{\partial F_{1,m}}{\partial F_{0,1}} = \frac{\partial V_{1}^{\pi}(x_{1}, F_{1})}{\partial F_{1,1}} \beta^{0,1}_{1}(Y) = \frac{\partial V_{1}^{\pi}(x_{1}, F_{1})}{\partial F_{1,1}} \frac{s_{1}}{F_{0,1}}. \tag{24}
\]

Pick a vector \( F_{1} \) at which \( V_{1}^{\pi}(x_{1}, F_{1}) \) is differentiable. It follows from Assumption 1(b) that

\[
\frac{\partial V_{1}^{\pi}(x_{1}, F_{1})}{\partial s_{1}} \Bigg|_{F_{1}=F_{1}^{\prime}} = \frac{\partial p(a_{1}^{\pi}(x_{1}, F_{1}^{\prime}), s_{1})}{\partial s_{1}} \Bigg|_{s_{1}=\pi_{1}} + \frac{\partial W_{1}^{\pi}(x_{1} - a_{1}^{\pi}(x_{1}, F_{1}^{\prime}), F_{1}^{\prime})}{\partial s_{1}} \Bigg|_{F_{1}=F_{1}^{\prime}}. \tag{25}
\]
The first term on the right hand side of (25) can be expressed as follows:

\[
\frac{\partial p(a_1^\pi(x_1, \overline{F}_1), s_1)}{\partial s_1} \bigg|_{s_1 = \pi_1} = \frac{\partial (\phi^1 s_1 + c^1) a_1^\pi(x_1, \overline{F}_1)}{\partial s_1} \bigg|_{s_1 = \pi_1} 1\{a_1^\pi(x_1, \overline{F}_1) < 0\} \\
+ \frac{\partial (\phi^W s_1 - c^W) a_1^\pi(x_1, \overline{F}_1)}{\partial s_1} \bigg|_{s_1 = \pi_1} 1\{a_1^\pi(x_1, \overline{F}_1) > 0\} \\
= \left( \phi^1 1\{a_1^\pi(x_1, \overline{F}_1) < 0\} + \phi^W 1\{a_1^\pi(x_1, \overline{F}_1) > 0\}\right) a_1^\pi(x_1, \overline{F}_1). \tag{26}
\]

As \( F'_1 \) does not depend on \( F_{1,1} \), the second term on the right hand side of (25) is zero:

\[
\frac{\partial W_1^\pi(x_1 - a_1^\pi(x_1, \overline{F}_1), F'_1)}{\partial s_1} \bigg|_{F'_1 = \overline{F}_1} = 0. \tag{27}
\]

It follows from (26) and (27) that (25) can be expressed as

\[
\frac{\partial V_1^\pi(x_1, F'_1)}{\partial F_{1,n}} \bigg|_{F_1 = \overline{F}_1} = \left( \phi^1 1\{a_1^\pi(x_1, \overline{F}_1) < 0\} + \phi^W 1\{a_1^\pi(x_1, \overline{F}_1) > 0\}\right) a_1^\pi(x_1, \overline{F}_1). \tag{28}
\]

Expressions (23), (24), and (28) imply that

\[
\Delta^\pi_{1,1}(0, x_1, F'_0) = \tilde{\delta}(0, T_1) \frac{\tilde{\delta}(0, T_1)}{F_{0,1}} \mathbb{E} \left[ \left( \phi^1 1\{a_1^\pi(x_1, F_1) < 0\} + \phi^W 1\{a_1^\pi(x_1, F_1) > 0\} \right) s_1 a_1^\pi(x_1, F_1) \right].
\]

Thus, the claimed property holds for \( m = 1 \) (recall that \( x_1 \equiv x_1^\pi \)).

The cases corresponding to \( m = 2, \ldots, N - 1 \) can be dealt with by recursively applying a logic similar to the case \( m = 1 \). \( \square \)

**Proof of Proposition 3 (Bounds on deltas).** We derive the inequality \( \Delta^\pi_{n,m}(t, x_n, G_n(t)) \leq \tilde{\delta}(t, T_m)\phi^W C^W \). The inequality \( \Delta^\pi_{n,m}(t, x_n, G_n(t)) \geq \tilde{\delta}(t, T_m)\phi^I C^I \) can be established in an analogous manner. It holds that

\[
\Delta^\pi_{n,m}(t, x_n, G_n(t)) = \frac{\tilde{\delta}(t, T_m)}{F(t, T_m)} \mathbb{E} \left[ \left( \phi^I A_m^\pi(x_n, F_m) 1\{A_m^\pi(x_n, F_m) < 0\} \right) s_m \right]
\]

where the first and last equalities follow from Proposition 2 and the property \( \mathbb{E}[s_m|F(t, T_m)] = F(t, T_m) \), respectively. \( \square \)
Proof of Proposition 4 (BH optimality). The objective function of model (13) is nonnegative for any value of \( q(t) \). Thus, setting \( q(t) = \Delta_n^\pi(t) \) optimally solves the version of this model obtained by replacing \( \Delta_n^{\pi,\phi}(t) \) with \( \Delta_n^\pi(t) \). □

Proof of Proposition 5 (CMVH positions). We omit the suffix \((t)\) for ease of exposition. Using the constraints in set \( Q_n^L \) to set to zero in the objective function the vector \( q^\pi_n \), rearranging, and ignoring constant terms yields the unconstrained model

\[
\min_{q^\pi_n} \quad (q^\pi_n)^T \, \text{diag} \left( G^\pi_n \right) \, \xi_{n-1,1}^\pi \, \text{diag} \left( G^\pi_n \right) \, q^\pi_n - 2 \, (q^\pi_n)^T \, \text{diag} \left( G^\pi_n \right) \, \xi_{n-1,1}^\pi \, \text{diag} \left( G^\pi_n \right) \, \Delta_n^\pi, \xi_{n}^\pi, \Delta_n^\pi, \xi_{n}^\pi
- 2 \, (q^\pi_n)^T \, \text{diag} \left( G^\pi_n \right) \, \xi_{n-1,1}^\pi \, \text{diag} \left( G^\pi_n \right) \, \Delta_n^\pi, \xi_{n}^\pi, \Delta_n^\pi, \xi_{n}^\pi.
\]

Because \( \xi_{n-1,1}^\pi \) is assumed positive definite, the necessary and sufficient optimality conditions for this model are

\[
\text{diag} \left( G^\pi_n \right) \, \xi_{n-1,1}^\pi \, \text{diag} \left( G^\pi_n \right) \, q^\pi_n - \text{diag} \left( G^\pi_n \right) \, \xi_{n-1,1}^\pi \, \text{diag} \left( G^\pi_n \right) \, \Delta_n^\pi, \xi_{n}^\pi, \Delta_n^\pi, \xi_{n}^\pi
- \text{diag} \left( G^\pi_n \right) \, \xi_{n-1,1}^\pi \, \text{diag} \left( G^\pi_n \right) \, \Delta_n^\pi, \xi_{n}^\pi, \Delta_n^\pi, \xi_{n}^\pi = 0.
\]

The optimal solution is thus

\[
q^\pi_n = \Delta_n^\pi, \xi_{n}^\pi + \text{diag}^{-1} \left( G^\pi_n \right) \left( \xi_{n-1,1}^\pi \right)^{-1} \, \xi_{n-1,1}^\pi \, \text{diag} \left( G^\pi_n \right) \, \Delta_n^\pi, \xi_{n}^\pi, \Delta_n^\pi, \xi_{n}^\pi. \square
\]

Proof of Proposition 6 (Approximate equivalence). We suppress the suffix \((t)\) to simplify the exposition. For a given \( K \) factor model let \( \Sigma_{n-1} \) be its corresponding matrix of factor loading coefficients and \( \xi_{n-1}^K \) the analogue for this model of the matrix \( \xi_{n-1}^{N-1} \). Interpret \( \xi_{n-1}^K (\Delta t) \) as a function of \( \Delta t \). For sufficiently small \( \Delta t \), taking a Taylor series approximation of each element of \( \xi_{n-1}^K (\Delta t) \) around 0 allows us to write \( \xi_{n-1}^K (\Delta t) \approx \Delta t \Sigma_{n-1} \left( \Sigma_{n-1} \right)^T \). The resulting version of model (15) is

\[
\min_{q \in \mathcal{Q}_n} \Delta t \, (q - \Delta_n^\pi)^T \, \text{diag} \left( G \right) \, \Sigma_{n-1} \, \left( \Sigma_{n-1} \right)^T \, \text{diag} \left( G \right) \, (q - \Delta_n^\pi) . \tag{29}
\]

We have \( \mathcal{H}_n \equiv \mathcal{L}_n \) and \( \mathcal{L}_n = \mathcal{L}_n \), so that

\[
\Sigma_{n-1} \equiv \begin{bmatrix} B_{n-1} & \mathbb{E}_{n-1} \\ \mathbb{E}_{n-1} & \mathbb{E}_{n-1} \end{bmatrix}
\]

and

\[
\Sigma_{n-1} \left( \Sigma_{n-1} \right)^T \equiv \begin{bmatrix} B_{n-1} B_{n-1}^T & B_{n-1} E_{n-1}^T \\ E_{n-1} B_{n-1}^T & E_{n-1} E_{n-1}^T \end{bmatrix}.
\]
Because the matrix $B_{n-1}B_{n-1}^T$ is positive definite, by the assumption on the matrix $B_{n-1}$, proceeding as in the proof of Proposition 5 yields the following optimality conditions for this model:

$$
\text{diag}(G^H_n) B_{n-1}B_{n-1}^T \text{diag}(G^H_n) q_{\pi,H_n} - \text{diag}(G^H_n) B_{n-1}B_{n-1}^T \text{diag}(G^H_n) \Delta_{n}^{\pi,H_n} - \text{diag}(G^H_n) B_{n-1}E_{n-1}^T \text{diag}(G^{\overline{\pi}}_n) \Delta_{n}^{\pi,\overline{\pi}_n} = 0.
$$

The solution to this system of linear inequalities is

$$
q_{\pi,H_n} = \Delta_{n}^{\pi,H_n} + \text{diag}^{-1}(G^H_n) \left( B_{n-1}^T \right)^{-1} B_{n-1} \text{diag}^{-1}(G^H_n) \text{diag}(G^H_n) B_{n-1}E_{n-1}^T \text{diag}(G^{\overline{\pi}}_n) \Delta_{n}^{\pi,\overline{\pi}_n}.
$$

**Proof of Lemma 1 (Characterization).** The claimed properties are true in stage $N-1$, because $V_{N-1}(x_{N-1},s_{N-1}) = \left( \phi^{W}W_{N-1} - c^{W} \right)^{+} (x_{N-1} \wedge C^{W})$, and $b_{N-1}(s_{N-1}) = 0$ and $\overline{b}_{N-1}(s_{N-1}) = \Phi_{1}\{ \phi^{W}W_{N-1} - c^{W} \leq 0 \}$. Make the induction hypothesis that these properties also hold in stages $n+1$, \ldots, $N-2$. Consider stage $n$. By definition, $W_{n}(x, F_{n}')$ is the discounted expected value of piecewise linear continuous functions, each with possible break points only in set $Q$, which implies that this function satisfies the same property. The quantities $b_{n}(F_{n})$ and $\overline{b}_{n}(F_{n})$ are optimal solutions to the following maximizations, respectively (see the proof of Secomandi 2010, Theorem 1):

$$
\max_{x_{n+1} \in X} W_{n}(x_{n+1}, F_{n}') - \left( \phi^{I}x_{n} + c^{I} \right) x_{n+1},
$$

$$
\max_{x_{n+1} \in X} W_{n}(x_{n+1}, F_{n}') - \left( \phi^{W}x_{n} - c^{W} \right) x_{n+1}.
$$

Thus, $b_{n}(F_{n})$ and $\overline{b}_{n}(F_{n})$ can be taken to belong to set $Q$. It follows that (recalling that $C^{I} < 0$)

$$
V_{n}(x_{n}, F_{n}) = \begin{cases} 
\left( \phi^{I}x_{n} + c^{I} \right) C^{I} + W_{n}(x_{n} - C^{I}, F_{n}'), & x_{n} \in [0, b_{n}(F_{n}) + C^{I}], \\
\left( \phi^{I}x_{n-1} + c^{I} \right) [x_{n} - b_{n}(F_{n})] + W_{n}(b_{n}(F_{n}), F_{n}'), & x_{n} \in [b_{n}(F_{n}) + C^{I}, b_{n}(F_{n})], \\
W_{n}(x_{n}, F_{n}'), & x_{n} \in [b_{n}(F_{n}), \overline{b}_{n}(F_{n})], \\
\left( \phi^{W}x_{n-1} - c^{W} \right) [x_{n} - \overline{b}_{n}(F_{n})] + W_{n}(\overline{b}_{n}(F_{n}), F_{n}'), & x_{n} \in (\overline{b}_{n}(F_{n}), \overline{b}_{n}(F_{n}) + C^{W}], \\
\left( \phi^{W}x_{n-1} - c^{W} \right) C^{W} + W_{n}(x_{n} - C^{W}, F_{n}'), & x_{n} \in (\overline{b}_{n}(F_{n}) + C^{W}, \overline{X}].
\end{cases}
$$
It is easy to check that this function is piecewise linear and continuous in $x_n$ with break points in set $\mathcal{Q}$ for each given $F_n$. Therefore, the claimed properties are true in stage $n$. By the principle of mathematical induction, they hold in every stage. \( \square \)

**Proof of Proposition 8 (Representation and Lipschitz continuity).**  (a) This part follows directly from part (b) of Lemma 1.

(b) The claimed property holds trivially in stage $N$ with $L_N(x) = 0$ for all $x \in \mathcal{X}$. Make the induction hypothesis that this property holds in stages $n + 1, \ldots, N - 1$. Consider stage $n$ and pick $F^1_n, F^2_n \in \mathbb{R}^{N-n}$. Define $C := |C^1| \lor C^W$ and $C' := (\phi^1 \lor \phi^W)C$. For each $a \in [-C, 0)$, it holds that

$$|p(a, s^2_n) - p(a, s^1_n)| = |(\phi^1 s^2_n + c^1)a - (\phi^1 s^1_n + c^1)a| \leq \phi^1|a||s^2_n - s^1_n| \leq \phi^1 C|s^2_n - s^1_n| \leq C'|s^2_n - s^1_n|.$$  

For $a = 0$, it holds that $|p(a, s^2_n) - p(a, s^1_n)| = 0 \leq \phi^W C|s^2_n - s^1_n| \leq C'|s^2_n - s^1_n|.$

For each $a \in (0, C]$, it holds that

$$|p(a, s^2_n) - p(a, s^1_n)| = |(\phi^W s^2_n - c^W)a - (\phi^W s^1_n - c^W)a| \leq \phi^W|a||s^2_n - s^1_n| \leq \phi^W C|s^2_n - s^1_n| \leq C'|s^2_n - s^1_n|.$$  

Thus, given $x \in \mathcal{X}$ and $a \in A(x) \subseteq [-C, C]$, we have

$$|p(a, s^2_n) - p(a, s^1_n)| \leq C'|s^2_n - s^1_n|. \tag{30}$$

Replacing $W_n(x_{n+1}, F^2_n) - W_n(x_{n+1}, F^1_n)$ with $\tilde{W}_n(x_{n+1}, F^2_n, F^1_n)$ for expositional convenience, it holds that

$$\left|\tilde{W}_n(x_{n+1}, F^2_n, F^1_n)\right| = \delta \left|\mathbb{E}\left[V_{n+1}(x_{n+1}, F^2_n \beta^{n,n+1}(Y)) - V_{n+1}(x_{n+1}, F^1_n \beta^{n,n+1}(Y))\right]\right| \leq \delta \mathbb{E}\left[\left|V_{n+1}(x_{n+1}, F^2_n \beta^{n,n+1}(Y)) - V_{n+1}(x_{n+1}, F^1_n \beta^{n,n+1}(Y))\right|\right]$$  

$$\leq \delta \mathbb{E}\left[L_{n+1}(x_{n+1}) \sum_{m=n+1}^N |F^2_{n,m} - F^1_{n,m}| \beta^{n,n+1}_m(Y)\right]$$

$$= \delta L_{n+1}(x_{n+1}) \sum_{m=n+1}^N |F^2_{n,m} - F^1_{n,m}| \mathbb{E}\left[\beta^{n,n+1}_m(Y)\right]$$

$$= \delta L_{n+1}(x_{n+1}) \sum_{m=n+1}^N |F^2_{n,m} - F^1_{n,m}|, \tag{31}$$

where the first inequality holds by the modulus inequality Resnick (1999, p. 128), the second inequality follows from the induction hypothesis, and the last equality follows from (20).
Given \( x \in \mathcal{X} \) and \( a \in \mathcal{A}(x) \), inequalities (30) and (31) imply that for all \( F_n^1, F_n^2 \in \mathbb{R}^N \) it holds that

\[
|v_n(x, a, F_n^2) - v_n(x, a, F_n^1)| = \left| p(a, s_n^2) - p(a, s_n^1) + W_n(x_n - a, F_n^2') - W_n(x_n - a, F_n^1') \right| \\
\leq \left| p(a, s_n^2) - p(a, s_n^1) \right| + \left| W_n(x_n - a, F_n^2') - W_n(x_n - a, F_n^1') \right| \\
\leq C'|s_n^2 - s_n^1| + \delta_{n+1}(x_n - a) \sum_{m=n+1}^N |F_{n,m}^2 - F_{n,m}^1| \\
\leq \{C' \vee \{\delta_{n+1}(x_n - a)\}\} \sum_{m=n}^N |F_{n,m}^2 - F_{n,m}^1|.
\]

Thus, the function \( v_n(x, a, F_n) \) is Lipschitz continuous in \( F_n \in \mathbb{R}^N \) for each given \( x \in \mathcal{X} \) and \( a \in \mathcal{A}(x) \).

Part (a) of this proposition and Dudley (2002, Proposition 11.2.2(a), p. 391) imply that the claimed property holds in stage \( n \). It follows from the principle of mathematical induction that the claimed property holds in every stage. \(\square\)

**Lemma 3.** Lemma 3 is needed in the proof of Proposition 9.

**Lemma 3** (Differentiability and unique optimal action). Under Assumption 2, for every \( n \in \mathcal{N} \), if \( V_n(x_n, F_n) \) is differentiable with respect to each element of \( F_n \) at a given futures curve \( \overline{F}_n \) for given \( x_n \in \mathcal{X} \), then at \( (x_n, \overline{F}_n) \) there is a unique optimal action, denoted as \( a_n^*(x_n, F_n) \), and

\[
\frac{\partial V_n(x_n, F_n)}{\partial F_{n,m}} \bigg|_{F_n=\overline{F}_n} = \frac{\partial v_n(x_n, a_n^*(x_n, \overline{F}_n), F_n)}{\partial F_{n,m}} \bigg|_{F_n=\overline{F}_n}, \forall m \in \mathcal{N}_n.
\]

**Proof.** Consider stage \( n \). Part (b) of Proposition 8 and Rademacher’s theorem imply that the function \( V_n(x_n, F_n) \) is differentiable almost everywhere in each element of \( F_n \) for each given \( x_n \in \mathcal{X} \). Fix \( x_n \) and pick \( \overline{F}_n \) such that \( V_n(x_n, F_n) \) is differentiable in each element of \( F_n \) at \( \overline{F}_n \). In particular, this means that \( \partial V_n(x_n, F_n)/\partial s_n \) exists at \( a_n^* \). Each function \( v_n(x_n, a, F_n) \) with \( a \in \mathcal{A}'(x_n) \) is linear in \( s_n \). This and part (a) of Proposition 8 imply that there is a unique optimal action at \( \overline{F}_n \), because otherwise \( \partial V_n(x_n, F_n)/\partial s_n \) would not exist at \( a_n^* \). This implies that \( V_n(x_n, F_n) \equiv v_n(x_n, a_n^*(x_n, F_n), F_n) \) in a neighborhood of \( F_n \). The assumed differentiability of \( V_n(x_n, F_n) \) in each element of \( F_n \) at \( \overline{F}_n \) implies that \( \partial V_n(x_n, F_n)/\partial F_{n,m} = \partial v_n(x_n, a_n^*(x_n, \overline{F}_n), F_n)/\partial F_{n,m} \) at \( \overline{F}_{n,m} \) for all \( m \in \mathcal{N}_n \). \(\square\)

**Proof of Proposition 9 (Pathwise deltas with an optimal policy).** Under Assumption 2, an optimal policy satisfies the conditions (a) and (b) stated in Assumption 1 by Proposition 8(b) and Lemma 3, respectively. Expression (19) then follows from Proposition 2. \(\square\)
### Table 7: Estimated standard errors of the HEM estimates reported in Table 2.

<table>
<thead>
<tr>
<th>K</th>
<th>DGNF = 3</th>
<th></th>
<th>DGNF = 23</th>
<th></th>
<th>DGNF = 3 plus Noise</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>NFH</td>
<td>FTFH</td>
<td>BH</td>
<td>NFH</td>
<td>FTFH</td>
<td>BH</td>
</tr>
<tr>
<td>1</td>
<td>1.81</td>
<td>1.85</td>
<td>0.43</td>
<td>1.89</td>
<td>1.89</td>
<td>0.35</td>
</tr>
<tr>
<td>2</td>
<td>7.56</td>
<td>0.67</td>
<td>0.27</td>
<td>62.05</td>
<td>0.52</td>
<td>0.22</td>
</tr>
<tr>
<td>3</td>
<td>0.40</td>
<td>0.25</td>
<td>0.24</td>
<td>167.01</td>
<td>0.26</td>
<td>0.18</td>
</tr>
<tr>
<td>5</td>
<td>6.47</td>
<td>0.24</td>
<td>0.24</td>
<td>913.91</td>
<td>0.21</td>
<td>0.18</td>
</tr>
<tr>
<td>10</td>
<td>0.30</td>
<td>0.24</td>
<td>0.24</td>
<td>391.49</td>
<td>0.17</td>
<td>0.17</td>
</tr>
<tr>
<td>15</td>
<td>0.33</td>
<td>0.24</td>
<td>0.24</td>
<td>304.62</td>
<td>0.18</td>
<td>0.17</td>
</tr>
<tr>
<td>20</td>
<td>0.26</td>
<td>0.24</td>
<td>0.24</td>
<td>49.82</td>
<td>0.18</td>
<td>0.18</td>
</tr>
<tr>
<td>23</td>
<td>0.24</td>
<td>0.24</td>
<td>0.24</td>
<td>0.18</td>
<td>0.18</td>
<td>0.18</td>
</tr>
</tbody>
</table>

### Table 8: Estimated standard errors of the HEM estimates reported in Table 4.

<table>
<thead>
<tr>
<th>K</th>
<th>L</th>
<th>DGNF = 3</th>
<th></th>
<th>DGNF = 23</th>
<th></th>
<th>DGNF = 3 plus Noise</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>3</td>
<td>5</td>
<td>10</td>
<td>3</td>
<td>5</td>
<td>10</td>
<td>3</td>
</tr>
<tr>
<td>1</td>
<td>0.98</td>
<td>0.85</td>
<td>0.34</td>
<td>1.44</td>
<td>1.13</td>
<td>0.44</td>
<td>1.05</td>
</tr>
<tr>
<td>2</td>
<td>1.01</td>
<td>0.76</td>
<td>0.25</td>
<td>1.39</td>
<td>1.03</td>
<td>0.35</td>
<td>0.98</td>
</tr>
<tr>
<td>3</td>
<td>1.03</td>
<td>0.75</td>
<td>0.25</td>
<td>1.39</td>
<td>1.02</td>
<td>0.34</td>
<td>0.95</td>
</tr>
<tr>
<td>5</td>
<td>1.04</td>
<td>0.74</td>
<td>0.25</td>
<td>1.39</td>
<td>1.00</td>
<td>0.34</td>
<td>0.95</td>
</tr>
<tr>
<td>10</td>
<td>1.05</td>
<td>0.75</td>
<td>0.25</td>
<td>1.40</td>
<td>1.00</td>
<td>0.34</td>
<td>0.94</td>
</tr>
<tr>
<td>15</td>
<td>1.05</td>
<td>0.75</td>
<td>0.25</td>
<td>1.40</td>
<td>1.00</td>
<td>0.34</td>
<td>0.94</td>
</tr>
<tr>
<td>20</td>
<td>1.05</td>
<td>0.75</td>
<td>0.25</td>
<td>1.40</td>
<td>1.00</td>
<td>0.34</td>
<td>0.94</td>
</tr>
<tr>
<td>23</td>
<td>1.05</td>
<td>0.75</td>
<td>0.25</td>
<td>1.40</td>
<td>1.00</td>
<td>0.34</td>
<td>0.94</td>
</tr>
</tbody>
</table>

### C. Standard Errors of the HEM and BHEM Estimates

Tables 7 and 8 include the standard errors (estimated by bootstrapping) of the HEM estimates in Tables 2 and 4, respectively. Tables 9 and 10 report the standard errors (also estimated by bootstrapping) of the BHEM estimates in Tables 5 and 6, respectively.
Table 9: Estimated standard errors of the BHEM estimates reported in Table 5.

<table>
<thead>
<tr>
<th></th>
<th>In-Sample</th>
<th>Out-of-Sample</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>NFH</td>
<td>FTFH</td>
</tr>
<tr>
<td>1</td>
<td>13.61</td>
<td>14.23</td>
</tr>
<tr>
<td>2</td>
<td>3360.35</td>
<td>0.23</td>
</tr>
<tr>
<td>3</td>
<td>7835.22</td>
<td>0.52</td>
</tr>
<tr>
<td>5</td>
<td>1771.76</td>
<td>0.72</td>
</tr>
<tr>
<td>10</td>
<td>1576.81</td>
<td>0.52</td>
</tr>
<tr>
<td>15</td>
<td>63.23</td>
<td>0.45</td>
</tr>
<tr>
<td>20</td>
<td>2.19</td>
<td>0.46</td>
</tr>
<tr>
<td>23</td>
<td>0.46</td>
<td>0.46</td>
</tr>
</tbody>
</table>

Table 10: Estimated standard errors of the BHEM estimates reported in Table 6.

<table>
<thead>
<tr>
<th></th>
<th>In-Sample</th>
<th>Out-of-Sample</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>L</td>
<td>L</td>
</tr>
<tr>
<td></td>
<td>K</td>
<td>3</td>
</tr>
<tr>
<td>1</td>
<td>13.94</td>
<td>8.63</td>
</tr>
<tr>
<td>2</td>
<td>12.67</td>
<td>7.65</td>
</tr>
<tr>
<td>3</td>
<td>10.01</td>
<td>5.71</td>
</tr>
<tr>
<td>5</td>
<td>9.04</td>
<td>5.03</td>
</tr>
<tr>
<td>10</td>
<td>8.49</td>
<td>4.63</td>
</tr>
<tr>
<td>15</td>
<td>8.47</td>
<td>4.62</td>
</tr>
<tr>
<td>20</td>
<td>8.49</td>
<td>4.63</td>
</tr>
<tr>
<td>23</td>
<td>8.48</td>
<td>4.62</td>
</tr>
</tbody>
</table>

OA-11
References


