MACRO STRATEGY

Surface Classification for Direction Predictions of the Foreign Exchange Implied Volatility Surface

June 8, 2020

Abstract

Predictions of the value of the FX implied volatility of an asset are commonly used as inputs to probabilistic models of profitability before expiration. However, depending on trading strategy, a metric that quantifies magnitudes of changes in implied volatility can be of supplemental value or of more interest than level prediction.

- Through transformation of the distribution of the observed implied volatilities, a predictable target representing the direction of change in the implied volatility of an asset can be generated.
- Correlation analysis of this target leads to significant dimensionality reduction with little loss of information.
- Cluster analysis of implied volatility surface data shows clear separation by overall change in the directions of surfaces.
- The clustering of surfaces shows evidence of Markov properties, leading to predictability of future values of cluster classification.
- Surface cluster classification shows predictability through historical, differenced, implied volatilities, interest rate and economic indicator data.
- The results of our directional forecasting models (i.e., Nested Markov Chain, Random Forest and Ensemble) show lower out-of-sample error when compared to deriving forecasts of the target using traditional value forecasts.

MetLife Investment Management

1. INTRODUCTION

Implied volatility (IV) is considered one of the most important variables for determining profitability in options trading. As a metric, IV indicates the market's view of the probability of movements in the price of a security. Predictions of the value of the IV of an asset are commonly used as inputs to probabilistic models of profitability before expiration. However, depending on the trading strategy, a metric that quantifies directions and magnitudes of changes in IV can be of supplemental value or of more interest than level prediction. In this study, the third paper of our Foreign Exchange (FX) Implied Volatility Forecasting research series, we proposed an approach to predicting the directional movement of IV based on surface dimension reduction. We compared the results of this approach to more standard approaches that show that the directional forecast approach can increase forecasting accuracy.

2. DATA

The data (source: Bloomberg) used in this study consisted of 75 points of the implied volatility surface for eight foreign exchange currencies from January 1, 2014 till June 11, 2019. The 75 points corresponded to five values of delta: 10 Call, 25 Call, At the Money, 25 Put and 10 Put, and fifteen values of expiry: 1, 2, and 3 weeks, 1, 2, 3, 4, 6, and 9 months, 1, 1.5, 2, 3, 4 and 5 years out. The feature data consisted of 2182 economic feature variables categorized by Commodity, Equity, Economic, Yield, Interest Rate, Currency Historical, Currency Implied Volatility, Currency Spot, EM Sector, and S&P 500 Sector data.

We generate the directional target as a categorical response with values $(-2, -1, 0, 1, 2)$ indicating large negative, small negative, zero, small positive, and large positive change in daily IV respectively for a pair, delta, expiry combination. The derivation of the historical values of the direction target is done by calculating Z-scores for each pair, delta, expiry combination across time and binning by quintile (i.e., -2, -1, 0, 1, 2).

3. METHODOLOGY

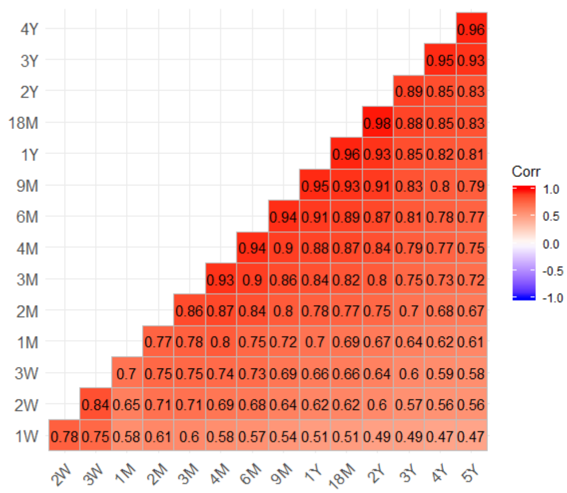
3.1. 3.1 Correlation Analysis

Correlation analysis, shown in Figure 1 and Figure 2, of these directional values lead to the conclusion that as there is little variation between deltas, dimension reduction by collapsing the delta dimension is appropriate. All considered models were therefore compared by their performance on modelling direction of at-the-money volatility and extrapolating to the other four delta values.

3.2. Cluster Analysis

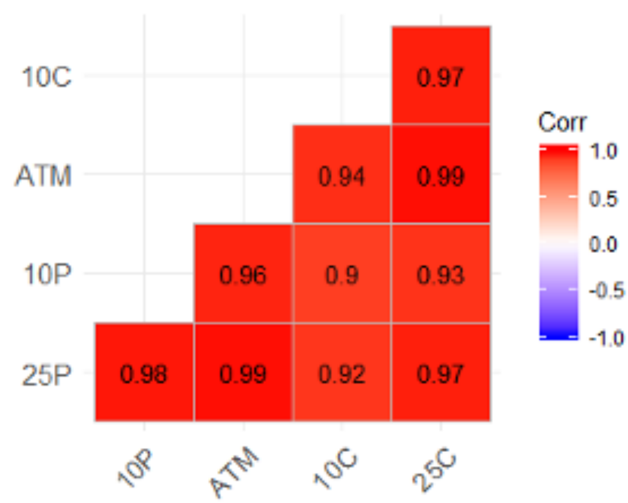
Further correlation analysis leads to the understanding that the majority of the variation in directional patterns across deltas and expiries could be explained by a set of simpler patterns. By treating the directional values of the 75 points as a

Figure 1: Correlations Analysis of Expiries



Data Source: MIM

Figure 2: Correlations Analysis of Deltas



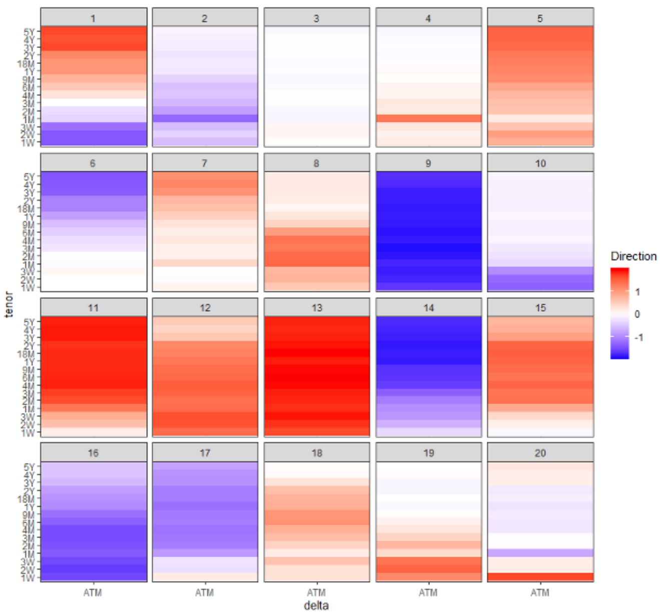
Data Source: MIM

surface, we find k surfaces that best explain the variation in actualized surfaces of directional values using k-means clustering. That is, for a given currency, given a set of observations $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n \in (-2,-1,0,1,2)^{15}$ where $\boldsymbol{y}_t$ is the 15-dimension directional surface corresponding to time t, we partition the n observations into k clusters $C_1, \ldots, C_k$ by minimizing the residual sum of squares:

$$\min \sum_{i=1}^{k} \mathrm{RSS}_i$$

where $RSS_i = \sum_{y \in C_k} \|\boldsymbol{y} - \boldsymbol{\mu}_k\|_2^2$ and $\boldsymbol{\mu}_k$ is the 15-dimensional mean vector of C_k. For illustration, the fitted clusters for $k = 20$ are shown in Figure 3. The problem framework is then reduced from a prediction of 75 categorical responses to the prediction of a single categorical response, the value of a cluster.

Figure 3: Fitted Clusters for k=20



Data Source: MIM

3.3. Prediction of Cluster Membership

Of the approaches considered for prediction of cluster classification, the highest performing were a nested Markov Chain model and multi-label classification model.

3.3.a Nested Markov Chain

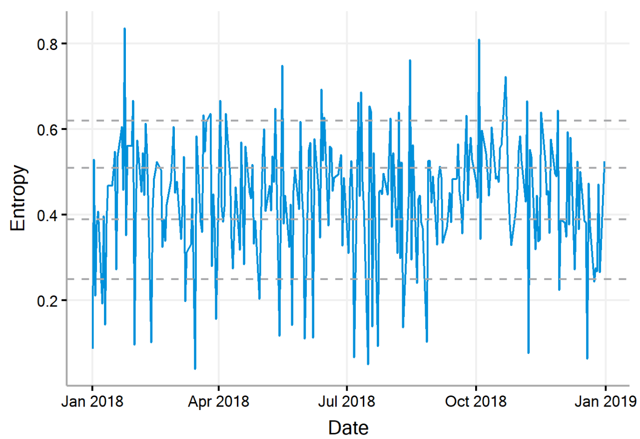
The assumption of the Markov property on the directional surface clusters is that for date $t+1$, the probability of a given value of cluster membership depends only on the value of cluster membership of date t. Under this assumption, we can estimate a $k \times k$ transition matrix from historical data and take the cluster prediction for date $t+1$ to be the average cluster weighted by the transition probabilities corresponding the date t. That is, for Markov transition matrix M, the one-day-out prediction of directional surface for date $t+1$ given cluster membership of j for date t based on this model is

$$\hat{S} = \sum_{i=1}^{k} M_{ij} C_i$$

This, unsurprisingly, was found to be an overly simplified assumption on the nature of cluster membership time series. However, it was found empirically that grouping historical data into u groups, which we will refer to as "economic states", and estimating u transition matrices, the cluster membership data more closely followed a Markov Chain model. These states were defined by clustering points in time with similar values of a value we call "entropy". Entropy here represents the variance in the correlations across currencies at the day level. To calculate the currency correlations, we vectorize the 75-point grid of implied volatility of a day for two currencies and compute the Pearson correlation coefficient of the two vectors. This allows for daily updates to correlation values in comparison to a more traditional approach of calculating a weighted average over a time window. The result of this calculation is $\frac{n \times (n-1)}{2}$ correlation values, $\{r_{ij}\}$. Entropy is then computed as the sample standard deviation of the correlation coefficients. States are then clustered using k-nearest neighbors.

The entropy time series and state classification for the year of 2018 is visualized in Figure 4.

Figure 4: Entropy of year 2018 with State classification rules



Data Source: MIM

After defining the u states, the $u \times u$ state transition matrix Θ is estimated from all data, and the u nested cluster transition matrices $M^{(v)}, v = 1, \ldots, u$ are estimated from the data in state v. The one-day-out prediction of directional surface for date $t+1$ given cluster membership of j and state membership of v for date t based on this model is

$$\widehat{M} = \sum_{v=1}^{u} \Theta_{iv} M_v$$

$$\hat{S} = \sum_{i=1}^{k} \widehat{M}_{ij} C_i$$

3.3.b Multi-label Classification Model

As described in Section 1, our dataset included feature data consisting of 2182 economic feature variables. The features were generated via PCA dimensionality reduction. Specifically, for each category of feature (see Section 2), we performed a separate PCA and took the components found to be correlated with the classification label. We also analyzed the importance of the generated features in the classification of the surface cluster. Using decrease in regression tree impurity as our metric, we found the most predictive components to be those generated from differenced values of IV, Interest Rate data, and Economic data. The Economic data components were largely functions of unemployment rates, capacity utilization, and purchasing manager indices. The components generated from IV data made up 30% of the total decrease in Gini, the interest rate data components made up 25%, and the economic data components made up 25%. The remaining 20% was distributed across the other feature sets. If we notate our data as $\{(x_i, C_i)\}_{i=1}^{n}$ where x_i is the feature vector for observation i and C_i is the value of cluster membership for observation i, we consider fitting a classification model to our data such that we can predict an unobserved C_{t+1} given a feature vector x_{t+1}. While there are many choices of classification algorithms, we choose to use a Random Forest model for this problem. We found that tree-based methods modelled the nonlinearity of the relationship of cluster membership and the economic covariate information well, but required a large value of depth. Indications of overfitting the training set was clear, so Random Forests were used to train multiple decision trees on subspaces of the feature space at the cost of slightly increased bias. Gini impurity was used as the splitting criterion. Given the classification probability predictions of the random forest $\hat{p}_i, i = 1, \ldots, k$, the prediction of directional surface based on this model is

$$\hat{S} = \sum_{i=1}^{k} \hat{p}_i \, C_{\boldsymbol{i}}$$

3.3.c Ensemble Model

When comparing the performance of all models, we found a discrepancy in "confidence". The most accurate models, that is, the models that had the highest rate of correct classifications, had larger spikes of error. This was to be expected, keeping in mind the ordinal nature of the response, but we sought a trade-off between accuracy and stability. We therefore constructed a simple ensemble model with the following logic:

1. If all models agree in sign of direction ("Down", "Zero" or "Up"), the ensembled prediction is that of the most confident model: the prediction vector with the largest 2-norm.

2. If models disagree in sign of direction, the ensembled prediction is the average of prediction vectors.

4. MODEL PERFORMANCE

We use several metrics for measuring model performance, each with a different view on error. Because the predictions are five dimensional and ordinal while the actual values are one-dimensional, special considerations need to be made when measuring performance. The first error metric, logloss, penalizes predictions that have large predicted probabilities of large negative directions when the actual direction was largely positive, for example. Logloss here is defined

$$\text{Loss} = -\frac{1}{4}\sum_{i=1}^{4}\left(\boldsymbol{\theta}_i \, \log \widehat{\boldsymbol{\theta}}_i + (\mathbf{1}-\boldsymbol{\theta}_i)\log(\mathbf{1}-\widehat{\boldsymbol{\theta}}_i)\right)$$

where

$$\boldsymbol{\theta} = \begin{pmatrix} I(y < \ -1) \\ I(y < 0) \\ I(y < 1) \\ I(y < 2) \end{pmatrix}, \widehat{\boldsymbol{\theta}} = \begin{pmatrix} \hat{y}_{-2} \\ \hat{y}_{-2} + \hat{y}_{-1} \\ \hat{y}_{-2} + \hat{y}_{-1} + \hat{y}_0 \\ 1 - \hat{y}_2 \end{pmatrix}$$

That is, $\boldsymbol{\theta}$ is a vector of accumulated actual indicators and $\widehat{\boldsymbol{\theta}}$ is a vector of accumulated predicted probabilities. Then the loss can be seen as the negative loglikelihood of a 4-dimensional binomial variable.

We also look at confusion matrices to get a sense for confidence of predictions under specific conditions. However, as noted, we must first transform our predictions to the scale of the actual directions. Given an expiry combination's predicted directional surfaces over time $\widehat{\boldsymbol{w}}_t$, we bin the values to quintiles and create confusion matrices like the one shown in Table 1.

Table 1: Example of Confusion Matrix

Predicted \ Actuals	-2	-1	0	1	2
Very Neg	7980	5072	5911	592	235
Neg	3851	4384	11238	1662	1160
Zero	5863	9196	34924	8248	6529
Pos	787	1019	6285	2657	3152
Very Pos	363	878	5945	3611	8963

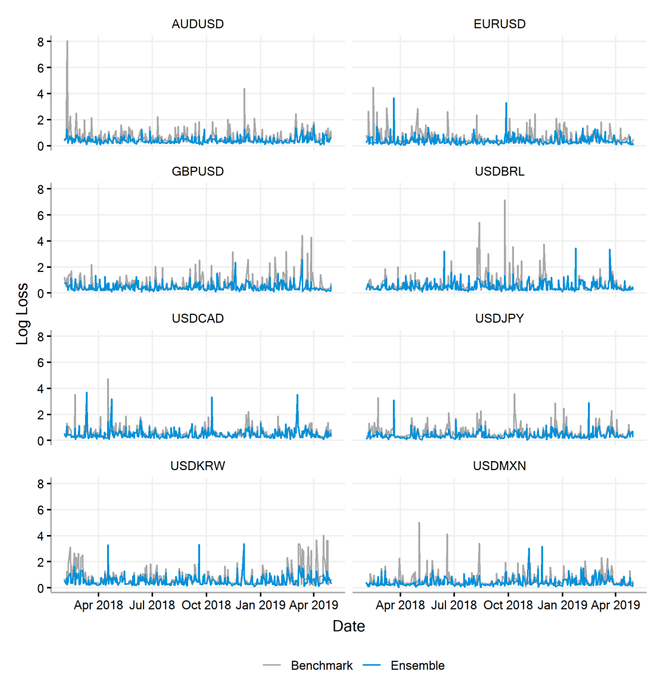
Data Source: MIM

OUT-OF-SAMPLE PERFORMANCE

We now observe the out-of-sample performance of the proposed methods compared to a benchmark. To establish a baseline model, we generate direction predictions derived from forecasts from a level model. Specifically, these baseline predictions are derived from IV level forecasts from an order-5 Vector-Autoregression (VAR) model [1,2]. The Z-scores were calculated for each pair, delta, expiry combination's forecast across time and direction forecasts were generated by quintile binning. The time frame of our dataset was January 1, 2014 to June 11, 2019. We took all data until January 1, 2018 as training data and iteratively fit the models and made one-day-out predictions over the remaining holdout data.

Figure 5 shows the logloss values of the benchmark and ensemble model predictions over time for the 3 Month ATM point for each pair. We can see that the benchmark model has instances of large errors, indicating highly confident predictions in the incorrect direction, while the ensemble odel has fewer and less egregious spikes across all pairs.

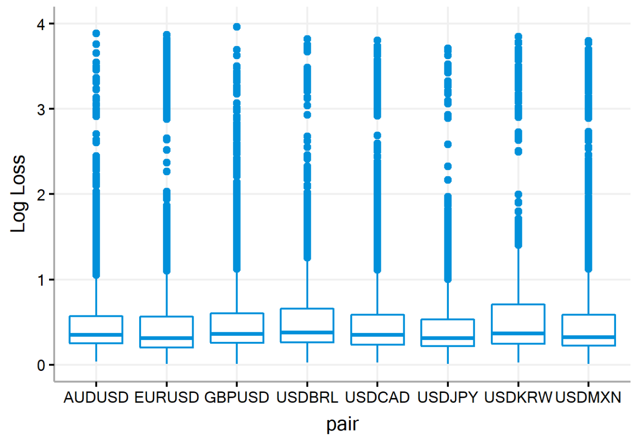
Figure 5: Logloss Spikes for all Pairs



Data Source: MIM

Figure 6 presents the distribution of logloss over the test data for all pairs. We can see that while some pairs have higher average values of error, there is no statistically significant evidence of bias over pairs.

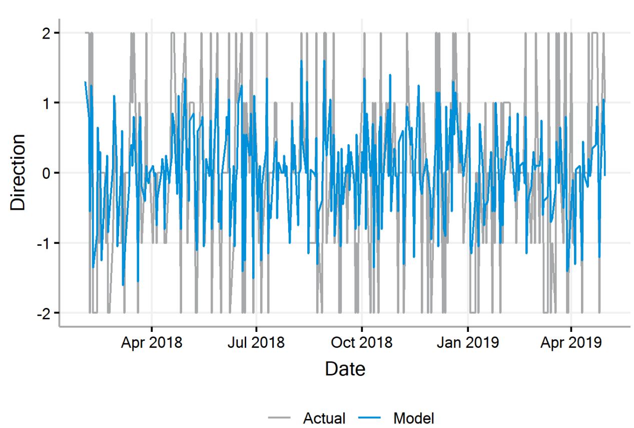
Figure 6: Logloss Distributions for all Pairs



Data Source: MIM

Figure 7 visualizes the magnitude of weighted movement predictions, that is, $\hat{S}$ from Section 1.3.b. of the ensemble model. We can see that there is a tendency towards smaller absolute weighted predictions. However, as discussed, this magnitude of prediction gives an ideal trade-off between confidence of forecasts and out of sample error.

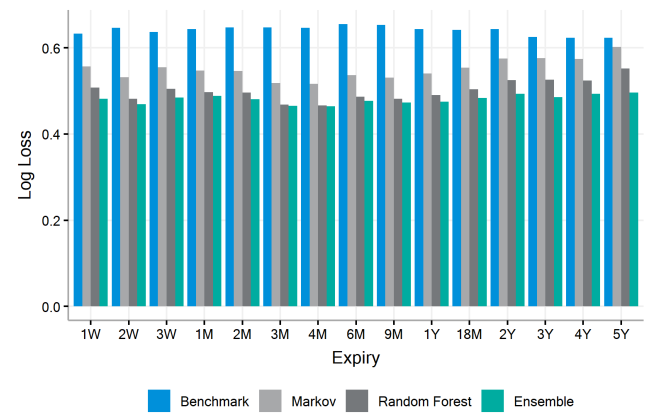
Figure 7: Magnitude of Model Predictions



Data Source: MIM

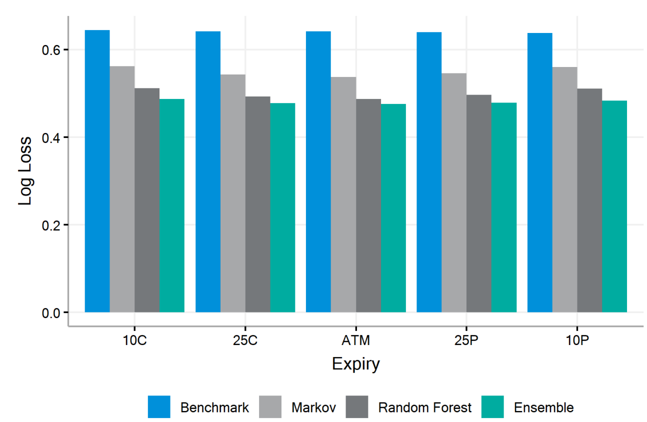
Figure 8 and 9 give the out-of-sample log-loss across all values of delta and expiry of the benchmark model, the proposed models using Markov Chain cluster prediction and the Random Forest cluster prediction, and the ensemble model. All proposed methods dominate the benchmark model across all expiries and deltas, and there is no clear bias to any of the points on the surface.

Figure 8: Logloss by Delta



Data Source: MIM

Figure 9: Logloss by Expiry



Data Source: MIM

Tables 2 and 3 give the confusion matrices of the out-of -sample predictions for the benchmark and the proposed model. We can see a clear increase in accuracy from the benchmark to the proposed model. We can also see significant marginal accuracy in the proposed model, especially in periods of large negative and positive movements.

Table 2: Confusion Matrix of Benchmark Model

Predicted \ Actuals	-2	-1	0	1	2
Very Neg	2971	3060	8489	2025	2299
Neg	2674	2978	9488	2449	2960
Zero	8255	9076	29820	7862	9290
Pos	2145	2411	7512	2161	2541
Very Pos	2799	3024	8994	2273	2949
Accuracy	0.15	0.14	0.46	0.13	0.15

Data Source: MIM

Table 3: Confusion Matrix of Ensemble Model

Predicted \ Actuals	-2	-1	0	1	2
Very Neg	7980	5072	5911	592	235
Neg	3851	4384	11238	1662	1160
Zero	5863	9196	34924	8248	6529
Pos	787	1019	6285	2657	3152
Very Pos	363	878	5945	3611	8963
Accuracy	0.42	0.21	0.54	0.16	0.45

Data Source: MIM

5. CONCLUSIONS

Implied volatility is essential to understanding the behaviour of asset prices and to implementing profitable trading strategies. In many cases, direction forecasting may be more reliable than value forecasting. Combining both value and direction forecasts, we can increase our confidence on market timing and making directional bets for Foreign Exchange options trading and hedging. We therefore performed a study using the implied volatility surfaces of eight foreign exchange currencies to develop and compare models for forecasting the direction of change in implied volatility. We found clear evidence of the predictability of the implied volatility direction of change as well as evidence that the proposed models outperformed a standard level-derived forecast.

REFERENCES

[1] Gong, Hanlu & Thavaneswaran, Aera & Singh, J. (2010). A Black- Scholes model with GARCH volatility. The Mathematical Scientist. 35.

[2] Bong Soo Lee and Doojin Ryu (2012). Stock Returns and Implied Volatility: A New VAR Approach. Economics Discussion Papers, No 2012-51, Kiel Institute for the World Economy. http://www.economics-ejournal.org/economics/discussionpapers/2012-51

Authors

Caleb Weaver, Ben Perryman, Jun Jiang, Alex Chien, Erez Meoded, Al Kirk

About MetLife Investment Management

MetLife Investment Management (MIM),[1] MetLife, Inc.'s (MetLife's) institutional investment management business, serves institutional investors by combining a client-centric approach with deep and long-established asset class expertise. Focused on managing Public Fixed Income, Private Capital and Real Estate assets, we aim to deliver strong, risk-adjusted returns by building tailored portfolio solutions. We listen first, strategize second, and collaborate constantly as we strive to meet clients' long-term investment objectives. Leveraging the broader resources and 150-year history of the MetLife enterprise helps provide us with deep expertise in navigating ever changing markets. We are institutional, but far from typical.

For more information, visit: **metlife.com/investmentmanagement**

[1] MetLife Investment Management ("MIM") is MetLife, Inc.'s institutional management business and the marketing name for the following affiliates that provide investment management services to MetLife's general account, separate accounts and/or unaffiliated/third party investors: Metropolitan Life Insurance Company, MetLife Investment Management, LLC, MetLife Investment Management Limited, MetLife Investments Limited, MetLife Investments Asia Limited, MetLife Latin America Asesorias e Inversiones Limitada, MetLife Asset Management Corp. (Japan), and MIM I LLC.

L0620004659[exp1120][All States], L0620004621[exp0622][All States], L0620004622[exp0622][All States], L0620004609[exp0622][All States]

MetLife Investment Management

One MetLife Way | Whippany, New Jersey 07981