GNMA II 30-YEAR PASS-THROUGH MBS PREPAYMENT ANALYSIS

SAM CHOI AND MIKE SCHUMACHER

GNMA II mortgage-backed securities were introduced in 1983, but took until early 1995 to reach the level of investor acceptance anticipated by the program's architects. In 1996, GNMA II monthly production surpassed the $2 billion mark.

Investor appetite for CMOs backed by GNMA II collateral has increased concurrently. In fact, over half of the CMOs backed by GNMA collateral that were issued in 1995–1996 used GNMA II pass-throughs. It is therefore important for fixed-income investors to analyze the prepayment patterns of GNMA II collateral.

Three features distinguish GNMA IIs from their GNMA I ancestors:

- GNMA II pools are allowed to be formed with mortgages from *multiple issuers*.
- GNMA IIs have a wide *weighted-average coupon* (WAC) *range*. Loan rates can be 50 to 150 basis points above the securitized net coupon versus 50 basis points in GNMA Is.
- GNMA II MBS have a *longer payment delay* than GNMA Is — fifty days versus forty-five days.

I. THEORETICAL FRAMEWORK

Prepayment models typically use the difference between gross WAC and prevailing market rates as a major determinant of projected speeds. The WAC dispersion and limited gross WAC data of GNMA IIs leave us unable to implement this approach. To remedy this problem, our method employs an indirect approach — we infer prepayment information from GNMA I data.

One might think that a GNMA II should always prepay faster than a comparable net coupon GNMA I.

SAM CHOI is assistant vice president at Smith Barney Mortgage Backed Securities Research in New York.

MIKE SCHUMACHER is director at Smith Barney Mortgage Backed Securities Research.

After all, the underlying mortgage holders are virtually identical, and GNMA IIs do have higher gross WACs than GNMA Is. GNMA IIs do generally prepay faster than GNMA Is, but the analysis is not quite that simple. The investor needs to consider the impact of WAC *level* and WAC *dispersion* on GNMA II prepayments.

Level

Level refers to the difference between the gross WAC on a GNMA II and the gross WAC on a GNMA I with the same pass-through rate (net coupon) as the GNMA II. Because the servicing spread on GNMA IIs can be 50–150 basis points, while the servicing spread on GNMA Is is 50 basis points, GNMA IIs have higher gross WACs than GNMA Is.

The effect of the high GNMA II gross WAC is almost always to increase prepayment speeds relative to GNMA I speeds. A high gross WAC results in decreased speeds only for pools that have become extremely burned out by experiencing multiple refinancing cycles.

Dispersion

Dispersion refers to the variance of the gross WACs in a GNMA II pool. Because the servicing spread on GNMA Is is always 50 basis points, GNMA I pools have zero WAC dispersion. However, because servicing spreads on GNMA IIs can range from 50 to 150 basis points, GNMA IIs do have WAC dispersion.

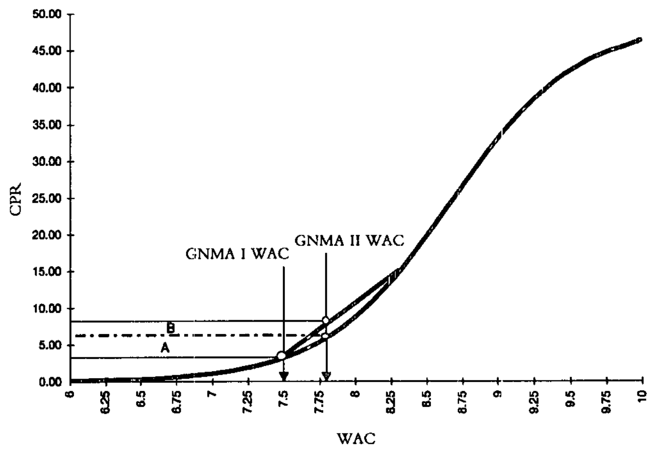
The impact of WAC dispersion on a pool's prepayment speed depends on whether the pool is a discount, current-coupon, or premium. In general, WAC dispersion causes prepayment speeds on discounts and current-coupons to increase relative to a similar WAC GNMA I pool, and speeds on premiums to decrease. WAC dispersion typically has a small effect on prepayment speeds of cusp coupons.

Level and Dispersion Effects

We diagram the WAC level and dispersion effects for an unseasoned discount GNMA in Exhibit 1. The net coupon on this pool is 7.00%. The WAC of the GNMA II is approximately 7.88%, and the WAC of a GNMA I would be 7.50%. The predicted prepayment rate of the GNMA I is approximately 3% conditional prepayment rate (CPR), and the predicted rate of the GNMA II is 8 CPR.

We isolate the effect of WAC level by computing the predicted prepayment speed for a pool with a 7.88% WAC but no WAC dispersion. In other words, the WAC effect is determined by moving along the GNMA I prepayment curve from the lowest point in the WAC range (i.e., the WAC that corresponds to a GNMA I with the same net coupon as the GNMA II) to the point corresponding to the GNMA II WAC. The WAC level effect is shown by area "A" in Exhibit 1, and the resulting speed is about 6 CPR.

EXHIBIT 1 ■ Impact of WAC Level and Dispersion on Prepayment Speeds



To analyze WAC dispersion, we follow these steps:

- Determine GNMA II WAC range. The low WAC equals N + 50 basis points, and the high WAC equals N + 2 × S – 50 basis points, where N is the net coupon and S is the actual servicing spread (the gross WAC minus the net coupon). Therefore, the midpoint of the range is N + S, the actual average gross WAC.
- Calculate predicted GNMA I speeds for pools corresponding to low WAC and high WAC (G1low and G1high).
- The predicted GNMA II speed is the average of G1low and G1high (G1average).
- Calculate predicted GNMA I speed for a pool corresponding to the actual WAC (G1actual).
- The WAC dispersion effect equals G1average minus G1actual.

The effect of WAC dispersion is shown by area "B" in Exhibit 1. The increase in prepayment speed resulting from WAC dispersion is approximately 2 CPR, and the resulting speed is 8 CPR.

As Exhibit 2 shows, WAC dispersion causes prepayment speeds to increase if the entire WAC range lies below the cusp of the S-curve (the point at which the slope of the tangent line reaches its peak) and causes speeds to fall if the WAC range is above the cusp. Equivalently, the WAC dispersion effect causes predicted prepayment rates to increase for coupons at which the GNMA I prepayment S-curve is convex, and causes predicted speeds to decline in areas where the S-curve is concave (negatively convex). The S-curve is convex below the cusp coupon, and concave above the cusp.

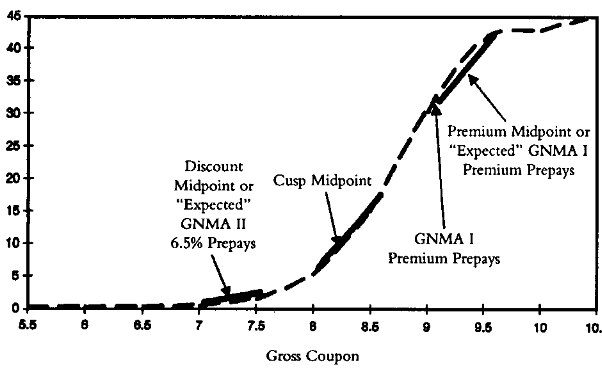
In Exhibit 1, the cusp corresponds to a WAC of approximately 8.75%. Therefore, WAC dispersion should cause predicted prepayment speeds to increase if the WAC range is below 8.75% (i.e., the net coupon is no higher than 8.25%), and should cause predicted speeds to fall if the entire WAC range is above 8.75%.

The predicted speed difference between GNMA II and GNMA I pass-throughs is a function of both WAC level and dispersion. The level effect is almost always positive, while we have seen that the impact of WAC dispersion is less straightforward. We expect most GNMA IIs to prepay faster than GNMA Is. Nonetheless, we recommend that investors use two submodels to incorporate the impact of WAC dispersion and capture the complexities of GNMA II prepayments.

II. EMPIRICAL RESULTS

While theoretical models can be elegant and interesting, a model's ability to predict empirical results is the ultimate indication of its value. We examine actual versus predicted prepayment behavior for GNMA II discounts, cusp, and high-premium issues, as well as consider the overall goodness of fit for regressions used to predict GNMA II prepayment speeds.

EXHIBIT 2 ■ Typical Refinancing S-Curve

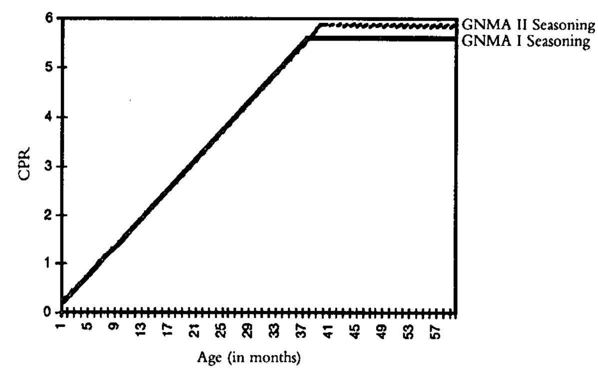


Discount GNMA II Prepayment Speeds

Our theoretical model suggests that prepayment speeds on discount GNMA IIs should be slightly faster than speeds on corresponding GNMA Is. Discount mortgages have experienced essentially no prepayment burnout, so the WAC level effect on discount prepayments is positive. As the S-curve is quite flat in the discount region, we expect the WAC level effect to be relatively small, however.

Discount mortgages are by definition below the cusp coupon. Therefore, as Exhibit 2 shows, the refinancing S-curve is convex in the discount area, and the WAC dispersion effect should be positive.

EXHIBIT 3 ■ Fitted Seasoning Curves for Discount GNMA I and II Pass-throughs



It appears that discount GNMA IIs do prepay minimally faster than GNMA Is. Exhibit 3 shows interpolated seasoning curves for discount GNMAs over the past eight years.[1] The curves suggest that discount GNMA IIs are very slightly faster than GNMA Is. The difference is within our statistical standard error and is significant only at the 15% level. Furthermore, the difference is not apparent until the pools are significantly seasoned. Exhibit 3 suggests the seasoning threshold is approximately three years.

Cusp and Premium Coupon Prepayment Behavior

GNMA II cusp coupons should be considerably faster than GNMA I cusps. By definition, the slope of the refinancing S-curve peaks at the cusp coupon. Therefore, a small increase in WAC results in a large increase in predicted prepayment speed. Consequently, the WAC level effect will be very significant.

As Exhibit 2 shows, the solid line connecting the high and low WACs in the cusp coupon GNMA II WAC range is virtually on top of the dotted line representing the GNMA I refinancing curve. Therefore, the prepayment speed corresponding to the midpoint of the solid line (the average of the high and low speeds) is very close to the speed matching the average WAC on the GNMA I refinancing curve. Consequently, the WAC dispersion effect should be close to zero.

In Exhibit 4, we report the mean CPR by coupon type for GNMA Is versus GNMA IIs. The statistics are based on historical generic data and include all bonds with an age of less than forty months. We include only newly issued bonds in an attempt to report prepayment rates in the absence of burnout effects. The time period of the analysis is January 1988–July 1996.

EXHIBIT 4 ■ Average CPR (weighted by amount outstanding) by Coupon Type

Mean CPR	GNMA I	GNMA II	GN II – GN I
Discount	3.6	4.1	0.5
Current	5.1	6.3	0.8
Cusp	30.2	31.8	1.6
Premium	41.6	44.0	2.4
Total No. of Obs.	1490	299	

To avoid any confusion, the discount category is defined as bonds with coupons 50 basis points or more below the current coupon. "Current" is self-explanatory. Cusps are bonds 50-100 basis points above the current coupon, and premiums are all coupons 100 or more basis points above the current coupon (not just bonds on the flat part of the S-curve).

Note also the dearth of historical data for GNMA IIs versus GNMA Is. Any model built using GNMA II data only would be less robust than a model that also uses any information that could be inferred from GNMA I data.

Goodness of Fit

The statistical fit for the GNMA II MBS S-curve in Exhibit 5 is not as accurate as the GNMA I MBS fit (the root mean squared error is 10% higher for GNMA IIs). We believe the difference in accuracy results from the small GNMA II data set and the gross WAC dispersion of GNMA IIs.

Ideally, all prepayment models would use gross WAC rather than net coupon as the prepayment reference rate. Unfortunately, GNMA II WAC data have been available for less than a year. Even if WAC data were available historically for many years, we would not get an accurate portrait of GNMA II prepayments unless we also captured the impact of WAC dispersion on prepayments.

One improvement we are implementing is to *model a GNMA II pool as two GNMA I pools* (akin to modeling multiple collateral groups within a CMO). Consider an example in which the securitized net coupon is 7.5%, and the gross weighted-average coupon of the securitized pool is 8.2% (in other words, the average servicing spread for this pool is 70 bp). Instead of modeling this as a GNMA II pool with an 8.2% WAC, we can evaluate it as two GNMA I subpools: 60% of the pool with an 8.0% WAC, and 40% of the pool with an 8.5% WAC (since $0.6 \times 8.0 + 0.4 \times 8.5$ is equal to 8.2). We can then model the GNMA II pool using the GNMA I prepayment S-curve and the formula:

GNMA II Prepayment Speed =
0.60(8.0% WAC GNMA I Speed) +
0.40(8.5% WAC GNMA I Speed)

Taking this approach, there are four parameters to be chosen:

1. The "GNMA I" coupon level on the lower-coupon subpool.
2. The "GNMA I" coupon level on the higher-coupon subpool.
3. The weight on the lower-coupon subpool.
4. The weight on the higher-coupon subpool.

EXHIBIT 5 ■ Fitted GNMA I and GNMA II Prepayment Curves

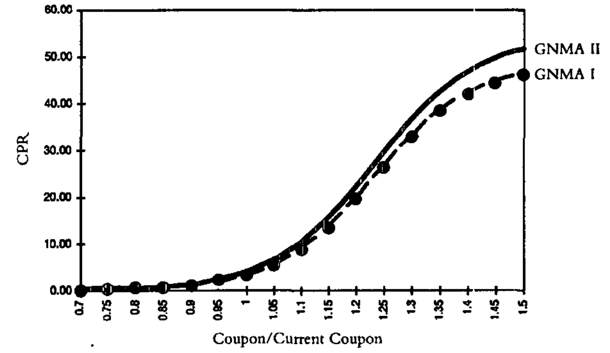


EXHIBIT 6 ■ GNMA II 7.5% MBS

Weight 1	Weight 2	Coupon 1	Coupon 2	Projection (CPR)	Dispersion (Projection – Base Case)	% Diff. from Base Case
1.00	0.00	7.50	0.00	3.61	N/A	N/A
1.00	0.00	7.75	0.00	4.60	0.00	0.0
0.50	0.50	7.50	8.00	4.85	0.25	5.4
0.75	0.25	7.50	8.50	5.41	0.81	17.6
0.85	0.15	7.50	9.00	6.24	1.64	35.7
0.25	0.75	7.65	7.78	4.60	0.00	0.0
0.50	0.50	7.65	7.85	4.64	0.04	1.0

In Exhibits 6 and 7, we calculate the predicted prepayments on a GNMA II 7.5% and GNMA II 9.5% MBS according to different combinations of the four parameters. The prepayments are based on a fitted GNMA I prepayment curve, and we assume that the current-coupon GNMA I MBS is 8.0%. Note that row 1 is the comparable GNMA I speed, and row 2 is considered the "base" or "no dispersion" case.

The choice of the four parameters is a mixture of art and science. Clearly, we want to choose a set of four parameters that fit the historical GNMA II data well. With limited GNMA II data, however, some degree of historical fit can be sacrificed to preserve model intuition. For example, even if there are no improvements in statistical fit from using different pairs of weights, the weight pairing that best reflects "perceived" WAC dispersion should be used (for example, closer to 50-50 than 100-0).

We can make some general comments from our research.

1. The "GNMA I" coupon level for the lower-coupon subpool should be 1-10 bp above the net coupon. For example, for a 7.5% GNMA II, the "GNMA I" coupon level for the lower-coupon subpool should be 7.51 – 7.60%.
2. The initial weights for premiums show a wider coupon dispersion than for discounts.

As Exhibit 4 shows, GNMA II discounts are only 0.44% CPR faster than their GNMA I counterparts. This small prepayment difference coincides with the minimal impact of WAC dispersion among discounts. GNMA II premiums are 2.44% CPR faster than their GNMA I counterparts. As Exhibit 7 shows, the difference between projected GNMA II and I prepayment rates decreases as WAC dispersion increases.

The tricky aspect of this approach is to track the relative weights of the two different GNMA I subpools as time progresses. Since the high-coupon subpool should initially prepay faster than the low-coupon pool, the weight of the high-coupon pool should diminish as

EXHIBIT 7 ■ GNMA II 9.5% MBS

Weight 1	Weight 2	Coupon 1	Coupon 2	Projection (CPR)	Dispersion (Projection – Base Case)	% Diff. from Base Case
1.00	0.00	9.50	0.00	23.25	N/A	N/A
1.00	0.00	9.75	0.00	33.04	0.00	0.0
0.50	0.50	9.50	10.00	32.45	–0.59	1.8
0.75	0.25	9.50	10.50	31.84	–1.20	3.6
0.85	0.15	9.50	11.00	30.99	–2.05	6.2
0.25	0.75	9.65	9.78	32.65	–0.39	1.2
0.50	0.50	9.65	9.85	32.68	–0.36	1.1

EXHIBIT 8 ■ Snapshot of GNMA II WACs by Coupon and Issue Year as of August 1996

Origination Year	Coupon	WAC	WAC – Coupon	Coupon	WAC	WAC – Coupon
1986	7.5	8.15	0.65	8.0	8.61	0.61
1987	7.5	8.16	0.66	8.0	8.64	0.64
1988	7.5	8.13	0.63	8.0	8.63	0.63
1989	7.5	8.11	0.61	8.0	8.60	0.60
1990	7.5	8.06	0.56	8.0	8.59	0.59
1991	7.5	8.11	0.61	8.0	8.59	0.59
1992	7.5	8.13	0.63	8.0	8.59	0.59
1993	7.5	8.07	0.57	8.0	8.58	0.58
1994	7.5	8.11	0.61	8.0	8.59	0.59

time progresses.[2] Thus, we expect to see the WAC of a GNMA II pool drift downward.

As an example, assume that we start with an equal number of mortgage holders in each of the subpools. Next, assume borrowers leave the high-coupon pool (prepay) at a 40% annual rate, while borrowers leave the low-coupon pool at a 20% annual rate. Then, after one year, we would expect the high-coupon pool to have approximately (1.0 – 0.40) or 60% of its original number of borrowers and the low-coupon pool to have approximately (1.0 – 0.20) or 80% of its original number of borrowers. High-coupon borrowers would therefore represent 43% of the total pool [60 / (60 + 80)], while the low-coupon borrowers would constitute 57% of the pool.

We would like to use a time series of actual GNMA II WAC information to test our theory that WACs drift downward. GNMA began providing WAC information only in mid-1996, however, so the time series is too short to be significant.

III. SUMMARY

The underlying mortgage holders in the GNMA I and GNMA II programs are extremely similar (mortgage holders conforming to GNMA criteria do not know if their loan has been included in a GNMA I or GNMA II pool). Our new method of modeling GNMA II prepayments is intuitive, and leads to improvements in statistical fit. Furthermore, differences between historical prepayment speeds on GNMA IIs and GNMA Is agree with our intuition. GNMA II discounts are slightly faster than GNMA Is, while speeds on unseasoned cusp and premium coupons are considerably faster than GNMA Is but tend to converge to GNMA I speeds as the pools season.

Only a sophisticated model that takes into account the coupon dispersion in the GNMA II collateral as well as the relative prepayment sensitivity within a given coupon group will model prepayments accurately over time. It is not enough simply to model GNMA II collateral as GNMA I collateral plus 25 basis points, because the servicing spread changes over time (Exhibit 8). Nor should one build a model using only the sparse GNMA II data — this not only reduces statistical fit accuracy, but also sacrifices model robustness.

ENDNOTES

[1]We interpolate using capitalization-weighted regression. A discount GNMA has a security coupon at least 50 bp less than the current-coupon MBS.

[2]If we want to predict prepayments on this pool two years out, the weights will have shifted, and it is more likely that there will be more of the pool in the 8.0% WAC level than the 8.5% WAC level. Thus, the effective WAC may be 8.1% instead of 8.25%.